Tungsten Oxide Memristive Devices for Neuromorphic Applications

by

Ting Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Wei Lu, Chair
Associate Professor Çagliyan Kurdak
Assistant Research Scientist Becky Lorenz Peterson
Assistant Professor Zhaohui Zhong

To my mom

"Perfect you in every good work for the doing of His will, doing in us that which is well pleasing in His sight through Jesus Christ….." Hebrews 13:21

ACKOWLEDGEMENTS

I would not have been who I am without His blessing and plentiful supply, especially the encouragement, nourishing, and cherishing from His loved ones. The Lord Jesus Christ is worthy of my highest praise!

Enormous gratitude is extended to my advisor, Prof. Wei Lu, who has given me the opportunity and constant support throughout my course of study. He has always been available and helpful, and has really set up a great example of blending smart mind and hard work into success. I have been granted exposures to many new experiences, from developing research ideas, preparing proposals, involving in collaborations, attending prestigious conferences, to mentoring students. I would like to thank my committee members for their useful discussions: Dr. Becky L. Peterson, Prof. Çagliyan Kurdak, and Prof. Zhaohui Zhong. I would also like to express my appreciation to my former and current group members: Sung Hyun Jo, Kuk-Hwan Kim, Wayne Fung, Siddharth Gaba, Patrick Sheridan, Lin Chen, Shinhyun Choi, Chao Du, and Dr. Yang for their helpful discussions and timely assistances, each of whom a brilliant individual.

I would also like to thank the Lurie Nanofabrication Facility (LNF) staff (and former staff) for their technical support: Sandrine Martin, Matt Oonk, Greg Allion, Brian VanDerElzen, Tony Sebastian, Russ Clifford, Vishva Ray, David Sebastian, Brian Armstrong, Terre Briggs, Nadine Wang, Tim Brock, and Ed Tang.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Inspired by the superb efficiency of biological systems, solid-state neural network systems have attracted much attention for their potential to learn and function in complex environments. Memristors, with several unique properties, are exceptional candidates for emulating artificial synapses and thus for building artificial neural networks. This thesis work explores the material properties, device characteristics, synaptic plasticity implementations, and CMOS integration of tungsten oxide (WO_X) nanoscale memristors, advancing this technology for neuromorphic applications.

Device fabrication, electrical studies, and material analyses of WO_X memristors are presented. Such devices offer simple fabrication, low-power, high-density, scalability, connectivity, and CMOS-compatibility, making them ideal candidates for neuromorphic applications. Bipolar analog resistive switching (RS) is observed as a result of oxygen vacancy (V_{OX}) migration within the WO_X film upon an applied electric field. Material characterizations suggest that the concentration and distribution of V_{OX} are directly related to the stoichiometry and morphology of the film. The switching characteristics are formulated into a set of memristive equations and further implemented in commercial simulation softwares, enabling simulations at the circuit level.

Based on the understanding of the switching dynamics of WO_X memristors, important synaptic functions are demonstrated in these electronic devices. Bipolar analog

RS implies that the memristor conductance (synaptic weight) can be potentiated (strengthened) and depressed (weakened) continuously by applying biases of opposite polarities. The inherent diffusion of V_{OX} results in limited retention in WO_X memristors. However, the retention can be enhanced by repetitively stimulating the memristor, known as the memory enhancement effect which mimics the transition of short-term memory to long-term memory in biological systems. Rate-dependent plasticity is studied by varying the interval between stimulations, and the results are analogous to post-tetanic potentiation and paired-pulse facilitation in neurobiology. Sliding threshold effect is also observed in WO_X memristors, indicating that conductance modulation is a dynamic process depending on the entire stimulation history, as stated in the BCM rule for biological systems. Heterosynaptic plasticity is also implemented through the design of a three-terminal memristor geometry. Finally, successful integration of WO_X memristors and with CMOS circuitry is carried out, facilitating the creation and characterization of hybrid memristor/CMOS neural network systems.

Chapter 1

Introduction

1.1 Background

The vision of building a computer version of the "brain", an entity capable of learning, remembering, thinking, reasoning, and making decisions has always been fascinating [1]. Modern computers are powerful systems based on the von Neumann architecture where computations take place sequentially between the processing unit (e.g. central processing unit, or CPU) and the storage unit (e.g. random access memory, or RAM). Historically, the performance improvements are mostly enabled by the faster components in each new generation. However, this architecture becomes less efficient with increasingly complex tasks even as the clock frequency reaches the order of GHz [2]. This effect is referred to as the “von Neumann bottleneck” reflecting the constraint that computing speed is ultimately limited by the traffic bandwidth between the processing and storage units. This bottleneck is now evident as we are more and more interested in computing tasks that are not well defined with large amount of data inputs (e.g. navigation of an autonomous vehicle in real traffic). On the contrary, animal and human brains operate at much lower frequencies but can perform these “fuzzy” logics extremely well with much less power consumption. The main difference is that information is parallel-processed in brains, i.e. processing and storage occur at the same time in the

same place, involving the corporate actions of large amounts of logic/memory elements. Inspired by biological systems, neuromorphic research [3] has generated a broad interest among scientists and engineers. One such example is a major initiative termed SyNAPSE (stands for Systems of Neuromorphic Adaptive Plastic Scalable Electronics) Program launched by the Defense Advanced Research Projects Agency (DARPA) in 2008 (Figure 1.1) [4]. The goal for this program is to build a neuromorphic machine that resembles the biological systems at the basic hardware level, requiring vigorous research efforts in hardware, architecture, simulation, and environment. Similar efforts such as the Blue Brain Project, the SpiNNaker Machine, and the Brain Scale S Neuromorphic Processors are being actively pursuit in Europe and worldwide [5].

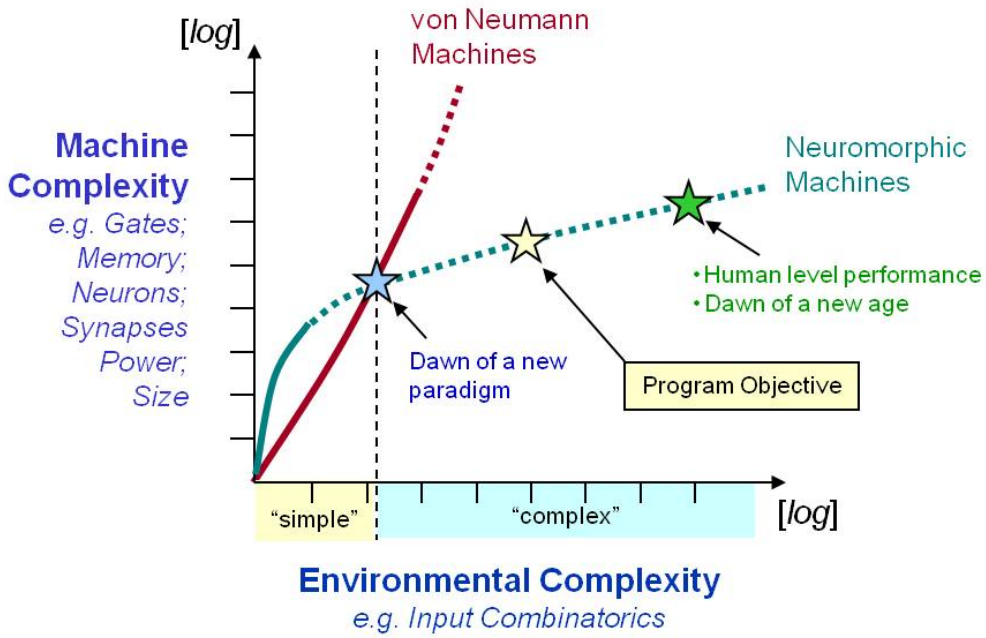


Figure 1.1 von Neumann Machines vs. neuromorphic machines. Image copied from DARPA's SyNAPSE Program description [4].

Advances in electronics, especially in complimentary metal-oxide-semiconductor (CMOS) and Flash memory technologies, have greatly impacted the entire world in the past few decades. The driving force is the down-scaling of device dimension, thus achieving higher speed, larger packing density and other favorable metrics [6]. Unfortunately, further scaling is not only facing manufacturing difficulties, such as lithography complexity and process controllability, but also soon approaching physical limits, for example, material properties and quantum effects [7]. To continue improving the performance and functionality of both technologies, novel device structures and physical mechanisms are being actively pursued by both industry and academic institutions [8]. One natural consequence of the emergent devices, however, is that they may no longer be well suited for conventional computing architectures but can instead enable new computing platforms such as neuromorphic computing discussed earlier.

One such example is resistive random-access memory (RRAM or ReRAM) devices based on two-terminal resistive switches which have drawn much attention for their use as non-volatile memory beyond Flash memory scaling [9-10]. Two-terminal RRAM devices have recently been categorized as "memristors" (memory+resistor) [11], regardless of materials or physical mechanisms utilized [12-13]. Strictly speaking, most two-terminal resistive switching (RS) devices do not follow the exact definition of memristors but fall in a broader category of memristive systems [14]. The shorter term "memristor" is used to refer to such devices throughout this dissertation simply out of convenience.

Memristors have been proposed for a wide range of applications such as non-volatile memory [10], synaptic computation [15], Boolean logic implementation [16],

signal processing [17], and circuit design [18]. In particular, the emulation of biological synapses with memristors for neuromorphic applications is of primary interest in this work [19-20]. In this chapter, fundamentals of memristors and their advantages for neuromorphic applications will be discussed.

1.2 Memristive Systems and Memristors

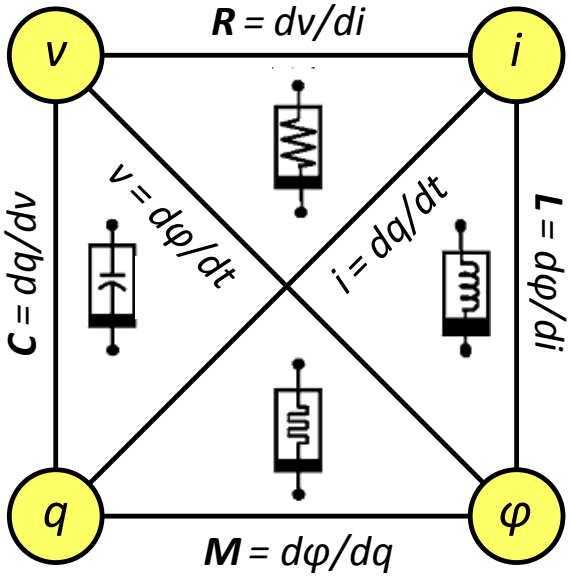


Figure 1.2 Functional relationships of the four physical variables: voltage (v), current (i), charge (q), and flux (φ), linked by the four circuit elements: resistor (R), capacitor (C), inductor (L), and memristor (M) [23].

The concept of a memristor was first proposed by Chua four decades ago [11]. His observation of the symmetry between the three fundamental circuit elements – resistor, capacitor, and inductor – led to the argument that there is a fourth element connecting the missing link between the charge and the flux to complete the relationships between the four physical variables: voltage, flux, current, and charge (Figure 1.2). This

element would behave as a nonlinear two-terminal resistor with memory, thus termed "memristor". Interestingly, Widrow had reported something very similar in 1960, albeit based on three-terminal devices, called “memistor” realized in a chemical device designed for adaptive systems [21].

Chua and Kang [12, 22] formulated the "memristive system" into a set of equations and generalized this definition to any two-terminal system that can be described by the theory when chosen appropriate variables and expressions. For a “voltage-controlled” memristor, this set of equations can be written as

$$I = G(w,v)v \tag{1.1}$$

$$\frac{dw}{dt} = f(w,v) \tag{1.2}$$

In Equation 1.1, *i* and *v* are related through a memductance (*G*) term. In a conventional resistive device, *G* is either fixed or a function of the instantaneous external input *v*. However, in a memristor, *G* not only depends on *v* but also depends on an internal state variable (*w*) which is not a simple function of *v*. Information about *w* is reflected in the dynamic equation (Equation 1.2) which describes the time derivative of *w*. These equations explain the memory effect since the state *w* is described by the time integral of Equation 1.2, and hence is history-dependent. Equation 1.2 is also called the “rate equation” as it governs the rate of change of *w*. The following discussion will be bound to the case of a memristor (memductance) in which only its resistance (conductance) is adjustable with a memory effect, and neglect other cases such as meminductors and memcapacitors [23].

A signature characteristic of memristors is the pinched hysteresis loop in the memristor *I*-*V* curve (Figure 1.3), indicating that memristors do not store energy as

capacitors and inductors do, but do have memory effects. Similar to resistors, capacitors and inductors, memristors are passive. Note that the memristor equations are simply mathematic expressions describing a nonlinear system, and do not infer any physical basis that would exhibit such characteristics. However, once a device has been verified to exhibit memristive effects and its state variable has been identified, these equations will be useful in describing the device behavior particularly in logic circuits, as will be discussed later.

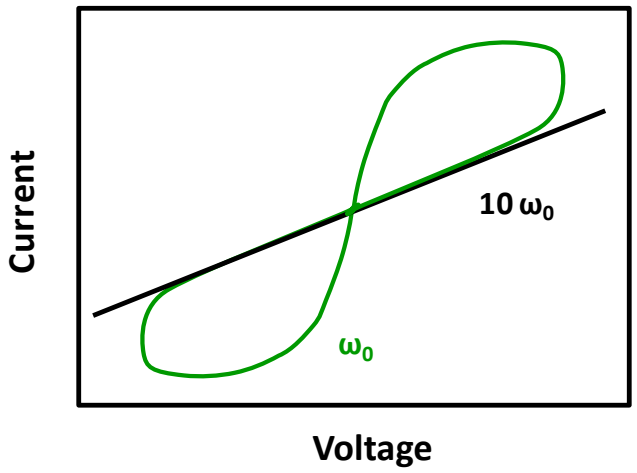


Figure 1.3 Pinched hysteresis loops for a typical memristor. The loop area shrinks as the sweeping frequency increases.

1.3 Nanoscale Solid-State Memristors

Studies on RS dated earlier than the theory the memristor, having a half-century history [25]. These studies mainly focused on the memory application potential of such devices so there was no drive to connect these devices with Chua's memristor which was proposed as a logic device. In 2008, HP Labs related the RS phenomena in titanium-oxide (TiO_2) thin-film devices to memristive effects and proposed that such devices can

indeed be explained by Chua's memristor model, and hence can be used for logic applications [13, 26, 27]. Besides TiO_2, RS effects have also been reported in other metal-oxides [28], amorphous semiconductors [29], Mott insulators [30], and organic materials [31]. Besides the large material variety, different structures (e.g. nanowire transistor in [32]) and substrates (e.g. plastic in [33]) are also available. Among the vast varieties mentioned, materials and processes that are compatible with standard semiconductor processing technology are especially interested for high-density integration based upon the CMOS platform.

1.4 Biological Synapses and Memristors as Synapses

After a brief introduction on memristors, we now discuss the potential of implementing them in neural computing architectures. Neurons and synapses together make up neural networks (Figure 1.4), which are the building blocks that empower humans to learn, think, and remember [34]. Human brain contains ~10^{10} neurons, each connecting to ~10^4 other ones. At each connection is a junction called synapse, so there are ~10^{14} synapses in a human brain. A key attribute of the brain's computing power is that the synapses are "plastic" – that is, the synaptic weight associated with each synapse can be modulated by signals (action potentials) generated from the two neurons joining the synapse and the new synaptic weight can be retained. The synaptic weight in turn determines the transmission between these the neurons. Thus, *synaptic plasticity* along with the very large *synaptic connectivity* empowers the efficient brain-based computing paradigm. Simply speaking, synaptic plasticity ensures learning ability while connectivity provides parallel processing capacity.

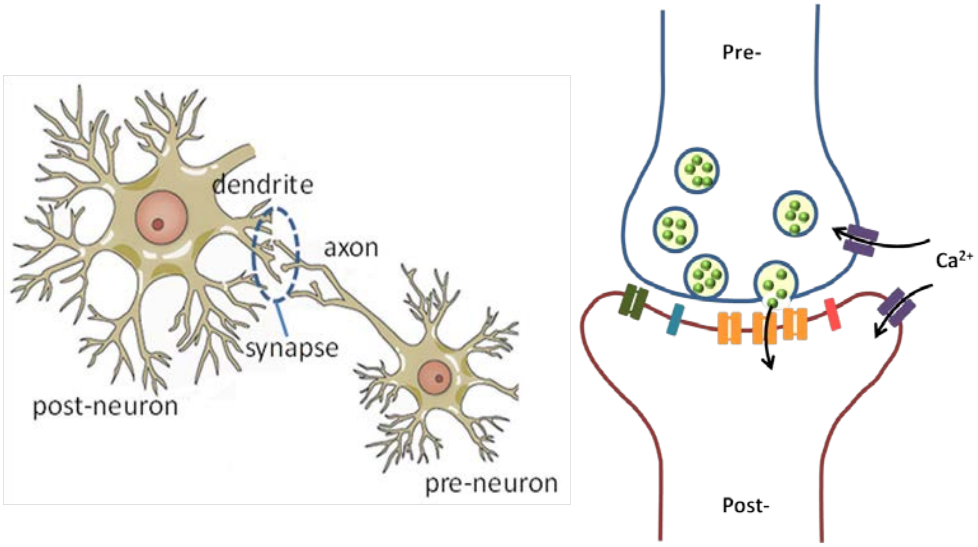


Figure 1.4 Illustrations of a neuron pair connected by a synapse (left) and neurotransmitters through a synapse junction (right).

So far, most attempts to implement neural networks took CMOS very large-scale integration (VLSI) technology as the backbone due to its process maturity and design flexibility. Digital circuits [35], analog circuits [36] and subthreshold low-power circuits [37] have been shown to emulate important synaptic learning rules while keeping the total power consumption and physical volume minimized. However, transistors are in nature three-terminal devices, implementation of large networks with massive plastic connections is extremely difficult with state-of-the-art interconnect technology. The difficulties come not only in designing and routing those interconnects but also the substantial power consumption associated with them. Another option is via software and algorithms at the expense of additional circuitry [38], for example, IBM's use of their Blue Gene/P supercomputer to simulate a cat brain. This work included higher level hierarchies than the hardware circuits previously listed, but could only simulate a cat brain at 1/83 speed of a real one at its best shot, and unsurprisingly, consumed enormous

resources (e.g. megawatts of power vs. ~20 W used by a human brain, volume, hardware, cost, etc.). No doubt, these approaches are impractical in the long run.

Considering the major drawbacks of different approaches mentioned above, the solution is to develop a nanoscale solid-state device capable of learning and remembering with ease of network connections. These properties are precisely the characteristics offered by memristors. Therefore, memristor devices are perhaps the most appropriate candidate for this application due to its memory, plasticity, connectivity, and power budget. In addition, such devices offer advantages such as structure simplicity, scalability, and process compatibility.

1.5 Integration and Neuromorphic Systems

Besides having memristors acting as synapses, one also needs devices to serve as neurons which are responsible for receiving and processing the information and firing action potentials (i.e. spiking) accordingly. CMOS circuits are suitable for this purpose for two reasons in particular: 1) CMOS circuits have been widely explored and are extremely versatile in performing different tasks so that various neuron models can be easily designed and implemented, and 2) since there are much fewer neurons than synapses, the challenges preventing CMOS circuits to act as synapses such as power, connectivity, and density are much more manageable here. Hence, to build a neuromorphic system, it is necessary to develop a solid-state nanoscale memristor device that uses CMOS-compatible materials and processes, so that the synaptic components can be seamlessly integrated with conventional CMOS circuits acting as neurons.

1.6 Organization of the Dissertation

In this chapter, the fundamentals of memristors and the advantages of using memristors for neuromorphic applications are introduced. Chapter 2 describes the fabrication and characterization of tungsten-oxide (WO_X) memristive devices, followed by their switching characteristics. Chapter 3 discusses the investigation of physical mechanisms responsible for resistive switching through designed material and electrical experiments. Based on the findings from previous studies, Chapter 4 provides a phenomenological model that can be implemented in circuit simulators (e.g. LTspice) and numerical computing programs (e.g. MATLAB), which facilitates efficient design for the future.

Chapter 5 and Chapter 6 present the key functionality of biological systems emulated in the WO_X memristors. Chapter 5 elaborates on the decay mechanism identified through retention studies , which suggest that the retention time of the memristor is not fixed but increases with its conductance state, known as the memory enhancement effect. Chapter 6 is dedicated to important synaptic learning rules such as rate-dependent plasticity, timing-dependent plasticity, the sliding threshold effect, and heterosynapticity. As the basis for memory and learning in brains, the successful demonstrations of these rules in memristors are significant.

Chapter 7 details the fabrication of high-density on-chip memristor arrays as well as the operation schemes of the peripheral circuitry for accessing the memristors. The successful integration of memristor arrays with existing CMOS technology enables the construction of functional neuromorphic networks. Finally, Chapter 8 concludes this thesis with a summary and suggestions for future work.

References

[1] M. Versace, B. Chandler, "The Brain of a New Machine," *IEEE Spectr*. **Dec**, 30, 2010.

[2] G. Snider, R. Amerson, D. Carter, H. Abdalla, M. S. Qureshi, J. Leveille, M. Versace, H. Ames, S. Patrick, B. Chandler, A. Gorchetchnikov, E. Mingolla, "From Synapses to Circuitry: Using Memristive Memory to Explore the Electronic Brain," *IEEE Computer* **11**, 37, 2011.

[3] C. Mead, "Neuromorphic Electronic Systems," *Proc. IEEE* **78**, 1629, 1990.

[4] http://www.darpa.mil/Our_Work/DSO/Programs/Systems_of_Neuromorphic_Adaptive_Plastic_Scalable_Electronics_(SYNAPSE).aspx

[5] http://www.artificialbrains.com/

[6] D. L. Critchlow, "MOSFET Scaling - The Driver of VLSI Technology," *Proc. IEEE* **87**, 659, 1999.

[7] Y. Taur, "CMOS Design Near the Limit of Scaling," *IBM J. Res. & Dev.* **46**, 213, 2002.

[8] T. Skotnicki, J. A. Hutchby, T. -J. King, H. -S. P. Wong, F. Boeuf, "The End of CMOS Scaling - Toward the Introduction of New Materials and Structural Changes to Improve MOSFET Performance," *IEEE Circuits Devices Mag.* **Jan-Feb**, 16, 2005.

[9] R. Waser, M. Aono, "Nanoionics-Based Resistive Switching Memories," *Nat. Mater.* **6**, 833, 2007.

[10] S. H. Jo, K. H. Kim, W. Lu, "High-Density Crossbar Arrays Based on a-Si Memristive Systems," *Nano Lett.* **9**, 870, 2009.

[11] L. O. Chua, "Memristor - The Missing Circuit Element," *IEEE Trans. Circuit Theory* **18**, 507, 1971.

[12] L. Chua, "Resistance Switching Memories Are Memristors," *Appl. Phys. A – Mater.* **102**, 765, 2011.

[13] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, "The Missing Memristor Found," *Nature* **453**, 80, 2008.

[14] B. Mouttet, "The Mythology of the Memristor," *IEEE ISCAS*, Paris, France, May 2010.

[15] G. S. Snider, "Cortical Computing with Memristive Nanodevices," *SciDAC Rev.* **Winter**, 58, 2008.

[16] J. Borghetti, G. S. Snider, P. J. Kuekes, J. J. Yang, D. R. Stewart, R. S. Williams, "'Memristive' Switches Enable 'Stateful' Logic Operations via Material Implication," *Nature* **464**, 873, 2010.

[17] L. O. Chua, L. Yang, "Cellular Neural Networks: Theory," *IEEE Trans. Circuits Syst.* **35**, 1257, 1988.

[18] Y. V. Pershin, M. Di Ventra, "Practical Approach to Programmable Analog Circuits with Memristors," *IEEE Trans. Circuits Syst. I, Regul. Pap.* **57**, 1857, 2010.

[19] S. H. Jo, T. Chang, I. Ebong, B. Bhavitavya, P. Mazumder, W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems, " *Nano Lett.* **10**, 1297, 2010.

[20] T. Chang, S. H. Jo, W. Lu, "Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor," *ACS Nano* **5**, 7669, 2011.

[21] B. Widrow, "An Adaptive 'Adaline' Neuron Using Chemical 'Memistors'," *Tech. Rep.* 1553-2, Stanford Electron. Labs., Stanford, CA, Oct. 17, 1960.

[22] L. O. Chua, S. M. Kang, "Memristive Devices and Systems," *Proc. IEEE* **64**, 209, 1976.

[23] M. Di Ventra, Y. V. Pershin, L. O. Chua, "Circuit Elements with Memory: Memristors, Memcapacitors and Meminductors," *Proc. IEEE* **97**, 1717, 2009.

[24] L. Chua, "Memristors: A New Nanoscale CNN Cell" in *Cellular Nanoscale Sensory Wave Computing*, C. Baatar, W. Porod, T. Roska, Ed.; Springer, 2010.

[25] J. G. Simmons, R. R. Verderber, "New Conduction and Reversible Memory Phenomena in Thin Insulating Films," *Proc. Roy. Soc. A* **301**, 77, 1967.

[26] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, R. S. Williams, "Memristive Switching Mechanism for Metal/Oxide/Metal Nanodevices," *Nat. Nanotechnol.* **3**, 429, 2008.

[27] D. B. Strukov, J. L. Borghetti, R. S. Williams, "Coupled Ionic and Electronic Transport Model for Thin-Film Semiconductor Memristive Behavior," *Small* **5**, 1058, 2009.

[28] A. Sawa, "Resistive Switching in Transition Metal Oxides," *Mater. Today* **11**, 28, 2008.

[29] S. H. Jo, K. H. Kim, W. Lu, "Programmable Resistance Switching in Nanoscale Two-Terminal Devices," *Nano Lett.* **9**, 496, 2009.

[30] L. Cario, V. Vaju, B. Corraze, V. Guiot, E. Janod, "Electric-Field-Induced Resistive Switching in a Family of Mott Insulators: Towards a New Class of RRAM Memories," *Adv. Mater.* **22**, 5193, 2010.

[31] D. I. Son, T. W. Kim, J. H. Shim, J. H. Jung, D. U. Lee, J. M. Lee, W. I. Park, W. K. Choi, "Flexible Organic Bistable Devices Based on Graphene Embedded in an Insulating Poly(methyl methacrylate) Polymer Layer," *Nano Lett.* **10**, 2441, 2010.

[32] J. I. Sohn, S. S. Choi, S. M. Morris, J. S. Bendall, H. J. Coles, W. K. Hong, G. Jo, T. Lee, M. E. Welland, "Novel Nonvolatile Memory with Multibit Storage Based on a ZnO Nanowire Transistor," *Nano Lett.* **10**, 4316, 2010.

[33] S. Kim, H. Y. Jeong, S. K. Kim, S. Y. Choi, K. J. Lee, "Flexible Memristive Memory Array on Plastic Substrates," *Nano Lett.* **11**, 5438, 2011.

[34] J. W. Rudy, *The Neurobiology of Learning and Memory*; Sinauer Associates, Inc. Publishers, 2008.

[35] G. Indiveri, E. Chicca, R. Douglas, "A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses with Spike-Timing Dependent Plasticity," *IEEE Trans. Neural Netw.* **17**, 211, 2006.

[36] P. Hasler, J. Dugger, "An Analog Floating-Gate Node for Supervised Learning," *IEEE Trans. Circuits Syst. I, Regul. Pap.* **52**, 834, 2005.

[37] G. Rachmuth, H. Z. Shouval, M. F. Bear, C. S. Poon, "A Biophysically-Based Neuromorphic Model of Spike Rate- and Timing-Dependent Plasticity," *Proc. Natl. Acad. Sci. USA* **108**, E1266, 2011.

[38] R. Ananthanarayanan, S. K. Esser, H. D. Simon, D. S. Modha, "The Cat is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses," *Proc. IEEE/ACM SC2009 Conf. High Performance Networking Computing*, Portland, OR, Nov. 2009.

Chapter 2

Tungsten Oxide-Based Memristors

2.1 Introduction

As mentioned in Chapter 1, resistive switching (RS) can be realized in various material systems utilizing different physical mechanisms. In order to take advantage of the advanced CMOS technology, using CMOS-compatible materials and processing techniques in memristor fabrication is essential. Throughout the progression of the integrated circuit (IC) technology, the industry has been very cautious in choosing metal materials and very faithful in staying confined to those selected metals as a result of research and manufacturing costs [1]. Initially, aluminum (Al) and tungsten (W) were used as interconnect and via metals, respectively. Beyond the 0.35 μm node, due to the superior conductivity and higher electromigration resistance, copper (Cu) was introduced with the aid of chemical mechanical polishing (CMP) processes. Besides Cu, a few other refractory metal materials serve as capping layers or as metal gates in recent generations, for example, titanium (Ti), titanium nitride (TiN), and tantalum nitride (TaN). In addition to material capability, if memristors were to be integrated with CMOS chips, back-end-of-line (BEOL) semiconductor processing further poses a temperature limit in subsequent steps to < 450 °C [2].

2.2 Material Overview

Considering the constraints listed above, tungsten oxide (WO_X) stands out exceptionally because on the one hand, W is CMOS-compatible and on the other, WO_X can be formed at rather low temperatures depending on deposition and formation techniques. Temperature as low as 300 °C is capable of oxidizing W into WO_X by plasma-enhanced oxidation, rapid thermal annealing (RTA) [3], or simple thermal oxidation [2, 4-5]; WO_X can be formed at room temperature by sputtering from a W target in an oxygen-containing environment [6] or by sputtering directly from a WO_3 target, [7-8]; WO_X can also be prepared by thermal evaporation of WO_3 powder [9], chemical vapor deposition (CVD) [10], pulsed-laser ablation [11], sol-gel technique [12], and so on.

Regarding applications, early understandings of WO_X were from metallurgy [13-14]. Later, WO_X has been studied extensively for its use in electrochromic devices [6, 9, 10, 12] and reducing gas sensors [7, 15]. More recently, WO_X was exploited in organic optoelectronics as transparent metal electrodes or charge-generation layers [16] and, certainly, in resistive switching memory devices [2-5, 8, 11]. It is an *n*-type semiconductor material with a bandgap between 2.6 eV and 3.4 eV, depending on its stoichiometry and morphology [9]. When WO_X is stoichiometric (i.e. $X = 3$), it is rather insulating; when it is substoichiometric (i.e. $X < 3$), the material is oxygen-deficient and becomes semiconducting or even metallic [17]; the electrical resistivity can range from 10^9 Ω-cm to 10^{-3} Ω-cm [9]. The crystal structure can also vary by a wide range, from amorphous to many crystal structures – monolithic, tetragonal, triclinic, hexagonal, orthorhombic, and cubic [14]. Optically, WO_X changes color from clear to yellow to blue

as X is decreased [2]. Upon the overwhelming information provided by chemists, material scientists, and device physicists, WO_X turns out to be an exceptionally versatile material to work with.

2.3 Device Structure and Fabrication

In a typical WO_X memristor, the layers form a metal-insulator-metal (MIM) structure, i.e. two electrodes sandwiching a switching layer, shown in Figure 2.1 [18]. To exclude the effect of electrode-WO_X interaction, inert metals (e.g. Pd, Pt, W, Au) were chosen. The fabrication process began with a thermally oxidized silicon substrate, followed by ~60 nm thick W deposition by DC sputtering at room temperature. Then the W film was patterned by electron-beam lithography (EBL) and reactive-ion etching (RIE) to form the bottom electrodes (BEs) and contact pads. Subsequently, the WO_X layer was formed by RTA in pure O_2 or plasma oxidation in O_2 plasma (in a PECVD chamber) above 350 °C. During oxidation, the top part of W was consumed (7 ~ 12 nm) and 20 ~ 35 nm of WO_X was formed. The thickness of WO_X film saturated after a certain thickness during the oxidation process due to finite oxygen diffusion at the oxidation temperature [2], for example, ~50 nm at 400 °C. Finally, the Pd top electrodes (TEs) and contact pads were formed by EBL and lift-off, followed by WO_X etch outside the cross-points using RIE to confine the active area to the crosspoints and also to expose the bottom contact pads. The final device had a Pd/WO_X/W structure with nanoscale junction sizes. The process flow for fabricating crossbar WO_X memristors is provided in Table 2.1.

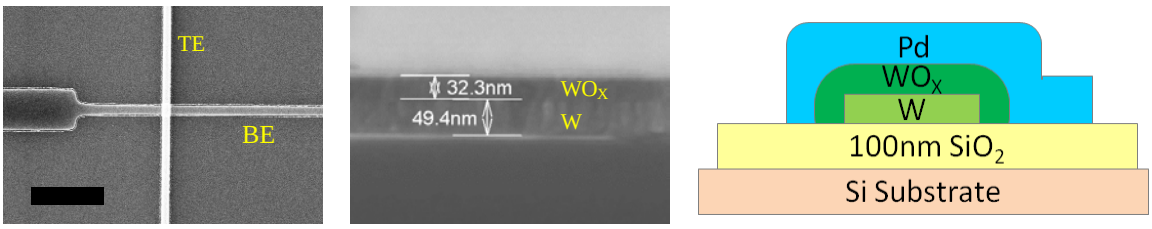


Figure 2.1 Scanning electron microscope (SEM) images and cross-section view of a WO_X memristor. (Left) SEM image of top view of a crossbar device with junction size ~ 130 nm x 130 nm (scale bar: 2 μm); (Center) SEM image of cross-section view of the WO_X layer oxidized from W in an RTA chamber at 375 °C for 1 min; and (Right) Cross-section schematic of the device structure.

Table 2.1 Process flow for stand-alone crossbar WO_X memristor fabrication.

Step	Process	Comment
	W deposition	60 nm W by DC Sputtering at room temperature (Kurt. J. Lesker LAB18)
	EBL for BE & contact pads	EBL (Raith 150) using PMMA A4 resist, MIBK:IPA = 1:3 developer (MicroChem)
	Ni evaporation & liftoff	40 nm Ni hard mask by e-beam evaporation (Cooke Evaporator), liftoff in acetone
	W etch	RIE using Cl_2/O_2 gas mixture with Ni as mask (LAM 9400)
	Ni removal	Removal of Ni by HCl:DI = 1:1 solution
	WO_X formation	Oxidation in O_2 ambient (JetFirst-150 RTP)
	EBL for TE & contact pads	EBL using PMMA A6 resist, MIBK:IPA = 1:3 developer
	Pd/Au evaporation & liftoff	80 nm Pd / 40 nm Au by e-beam evaporation, liftoff in acetone
	WO_X etch	RIE using SF_6/C_4F_8 gas mixture

This is the simplest structure since it takes as few as two lithography steps and does not require any planarization processes. Photolithography can also be used although larger devices potentially increase the leakage current [4]. To achieve high-density arrays

and integration with CMOS chips, spacers and/or spin-on glass (SOG) will be necessary, further complicating the fabrication. The topic on integration will be covered in Chapter 7.

2.4 Analog Switching Characterization

2.4.1 Switching Characteristics

In the designed testing scheme, BE is always grounded with voltage signals fed into TE. Current is read through a current amplifier at the BE terminal (Figure 2.2). RS in WO_X memristors is bipolar, meaning that the voltage polarities for increasing/decreasing device resistance are opposite, in contrast to unipolar devices where the resistance can be increased/decreased with single-polarity signals by adjusting current compliance (CC) levels. The Pd/WO_X/W memristors become more conductive (resistive) when applied with a positive (negative) bias, as shown in Figure 2.3. The conductance (or resistance) change is attributed to the re-distribution of oxygen vacancies (V_{OX}) within the WO_X film, creating (removing) conductive regions between the two electrodes. Detailed studies on material analyses and switching mechanisms will be provided in Chapter 3. Here switching is “analog,” referring to the incremental modulation of the device conductance, as opposed to “digital” switching where the device switches between two dominant (ON and OFF) states. In neurobiology, the conductance-modifying processes of synapses are called *potentiation* and *depression*, denoting that the synaptic weight is being strengthened and weakened, respectively [19]. Moreover, this synaptic conductance change is analog and rarely saturates at boundary values. It is noteworthy that voltage as low as 1 V is sufficient for both potentiation and depression, suggesting the opportunity

for low-voltage operation. In addition, current level below 50 μA guarantees low power dissipation and drivability at < 90 nm generations [20].

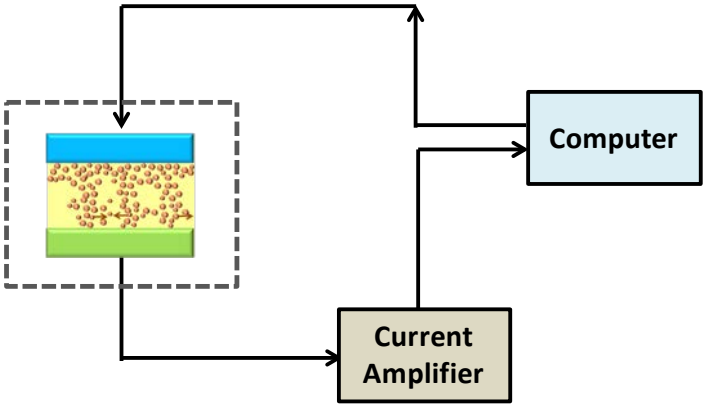


Figure 2.2 Measurement setup for characterization of WO_X memristors. The computer used LabView or LabWindows programs to generate input signals and collect output signals from the current amplifier. The device under test (DUT) was shielded in a probe station chamber at atmosphere pressure.

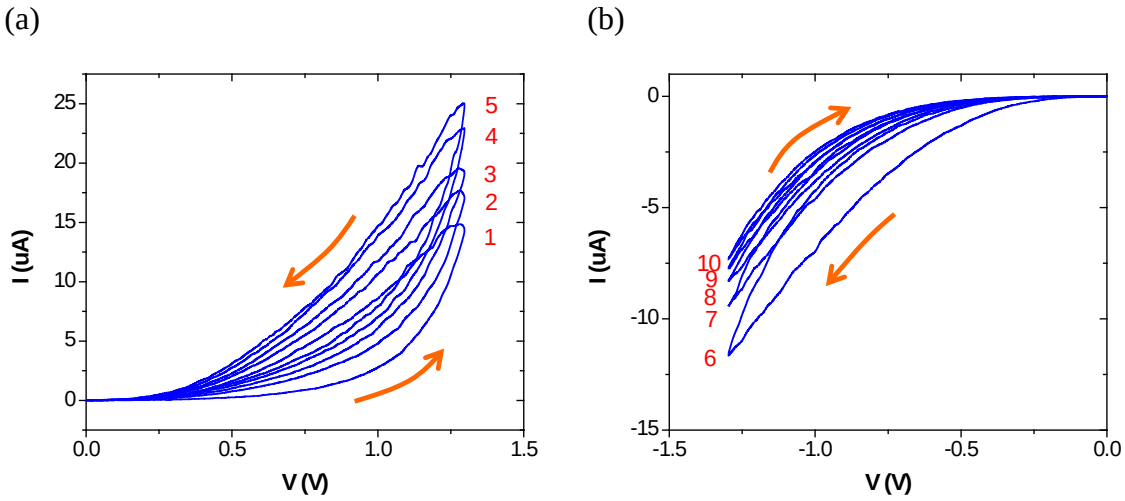


Figure 2.3 Bipolar resistive switching in a WO_X memristor. (a) Positive bias sweeps (1-5) continually enhance the device conductance while (b) negative bias sweeps (6-10) continually reduce the device conductance. Switching follows an analog fashion. Sweep rate was 2 V/sec for both polarities.

In Figure 2.3, the *I-V* hysteresis loops are the direct consequence of the memory effect. Additionally, the overlapping of these loops indicates that the device has limited retention, i.e. the conductance state of the device cannot be maintained for long time without refreshing signals and would naturally drift towards the less conductive (or more resistive) states. More discussions on the limited retention can be found in Section 5.2.

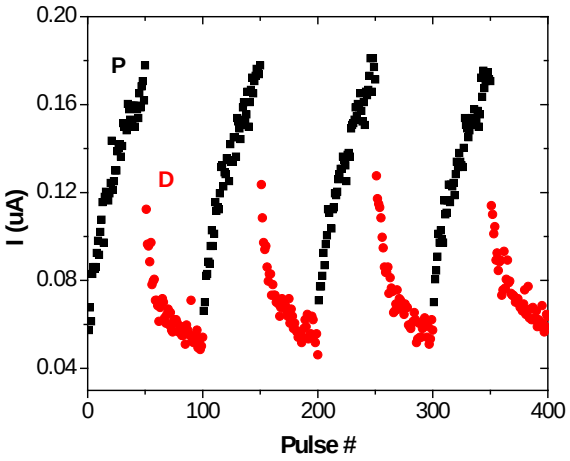


Figure 2.4 Pulse operation of a WO_X memristor. The data show the read current measured with a read pulse (0.4 V, 3 ms) following each potentiation (P: black) and depression (D: red) pulse. P: +1.3 V, 25 μs, D: -1.3 V, 25 μs.

The WO_X devices can also be operated with voltage pulses, which are more realistic signals than DC in ICs (Figure 2.4). Switching polarity is the same as that in DC operation. Pulses as short as a few microseconds were able to gradually change the device conductance. In fact, others have shown resistive switching in tantalum oxide (Ta_2O_5) with voltage pulses on the order of picoseconds consuming ~ pJ of power per switching event [21]. To read out the state of the device, read pulses at low voltages were used to avoid disturbing the instantaneous state. Measurements were performed through

custom-written LabView or LabWindows programs. The read pulse duration in these measurements was at least a few milliseconds, constrained by the data acquisition speed.

2.4.2 Endurance and Retention

If a memristor starts to lose plasticity or experience fatigue after several operations, it obviously would not be suitable for adaptive systems which are constantly changing according to the surroundings. Fortunately, the WO_X memristors always show robust endurance behavior. Figure 2.5 shows the response of the device to the same pulse train repeated at fresh, after 30,000 cycles, and after 100,000 cycles. A longest record of ~1.5 x 10^8 cycles has also been obtained [22]. No obvious degradation was observed, and the devices are believed to be able to undergo many more cycles, as a 10^{10}-cycle endurance has been reported in TaO_X-based devices [23].

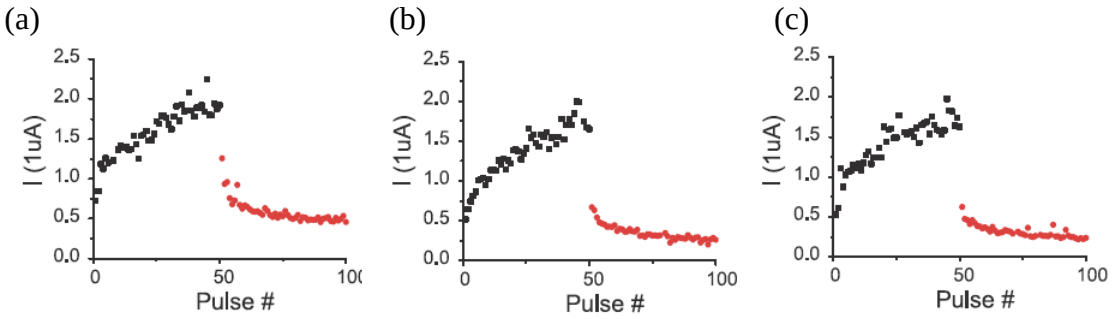


Figure 2.5 Endurance characteristics. P: +1.7 V, 80 μs, D: -1.7 V, 80 μs, and read: 0.5 V, 3 ms. P/D responses were measured at (a) fresh, (b) after 30K cycles, and (c) after 100K cycles.

As seen in the overlapping loops in Figure 2.3, due to the very mobile nature of V_{OX}, they tend to diffuse after the removal of the power supply and will settle down to equilibrium. The lower-conductance states are always more favorable than the higher-

conductance states, the same trend was observed in WO_X and TiO_2 memristors [24]. The retention time of such analog memristors ranges from a few seconds to a few hours, depending on the conductance states. A comprehensive study on retention and state decay will be presented in Chapter 5.

2.4.3 Switching Uniformity across Devices

Characterization of various aspects of WO_X memristors has provided much knowledge about the general behaviors of these devices. To utilize the memory and processing capabilities of memristors in large networks, it is necessary to study the statistical distribution of conducting states of single memristors as well as individual memristors in an array. The operating voltage, conducting current, retention time, and responsiveness to stimulations should each have a reasonably tight distribution. It has been shown by us that the fluctuations during DC sweeps from 7 different devices are within 2% [25]. Figure 2.6 plots 15 responses of a single device to an identical pulse train. Similarly, Figure 2.7 plots the responses of 8 different devices to the same pulse train. As expected, some fluctuations are evident during pulse operations but overall good switching uniformity can still be obtained. It is worth noting that synaptic connections in brains also constantly experience fluctuations and noises, nevertheless can process information correctly owing to their fault-tolerant architecture [26]. Therefore, the reasonable switching uniformity and repeatability among and within memristors demonstrated here should enable potential implementation in larger-scale networks.

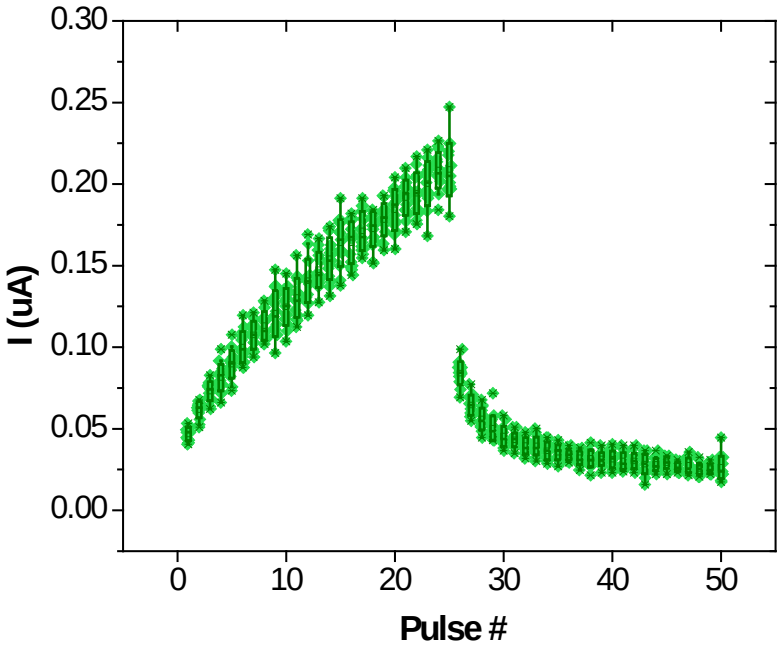


Figure 2.6 Switching uniformity from a single device. The dots show data from 15 separate pulse operations, and the boxes indicate the standard deviation (SD). The pulse train consisted of 25 potentiating pulses (+1.3 V, 200 μs) followed by 25 depressing pulses (-1.3 V, 200 μs), and the data were read by read pulses (+0.4 V, 3 ms) after each potentiating or depressing pulse. Device size: 350 nm x 350 nm.

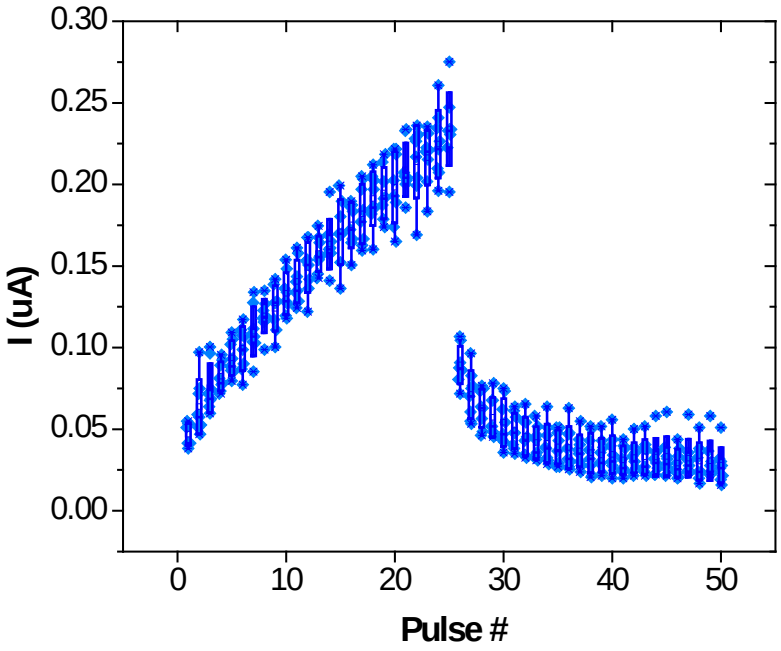


Figure 2.7 Switching uniformity from 8 different devices. Testing conditions and device specifications were identical to that of Figure 2.6.

2.4.4 Relation to Digital Switching Devices

Although WO_X memristors normally operate as analog switching devices, it is possible to turn them into digital switching devices through controlled programming, even after fabrication is completed. By constantly applying strong positive DC sweeps or pulses, the large electric field along with the current-induced heat within the nanoscale junction would cause pronounced change to the V_{OX} distribution and possibly involve structural changes of the WO_X film too. Figure 2.8 shows an example of an analog device that had become digital after many strong positive pulses at higher CC; potentiating and depressing pulses in (a) cause continuous modulation in the device conductance while the same pulses in (b) switch the device to the maximum and minimum conductance levels in step-like fashion. Whenever analog-to-digital switching is taking place, the retention is also significantly extended (Figure 2.8 (c) and (d)) because the structural changes in WOx associated with the digital switching is much more robust than the changes in V_{OX} distribution associated with analog switching. Indeed, there have been studies on using WO_X devices as digital RRAM [2-5] because WO_X can be conveniently formed on top of existing W plugs in CMOS processes. Note that due to the severe change in the film, the analog-to-digital transition is typically not reversible.

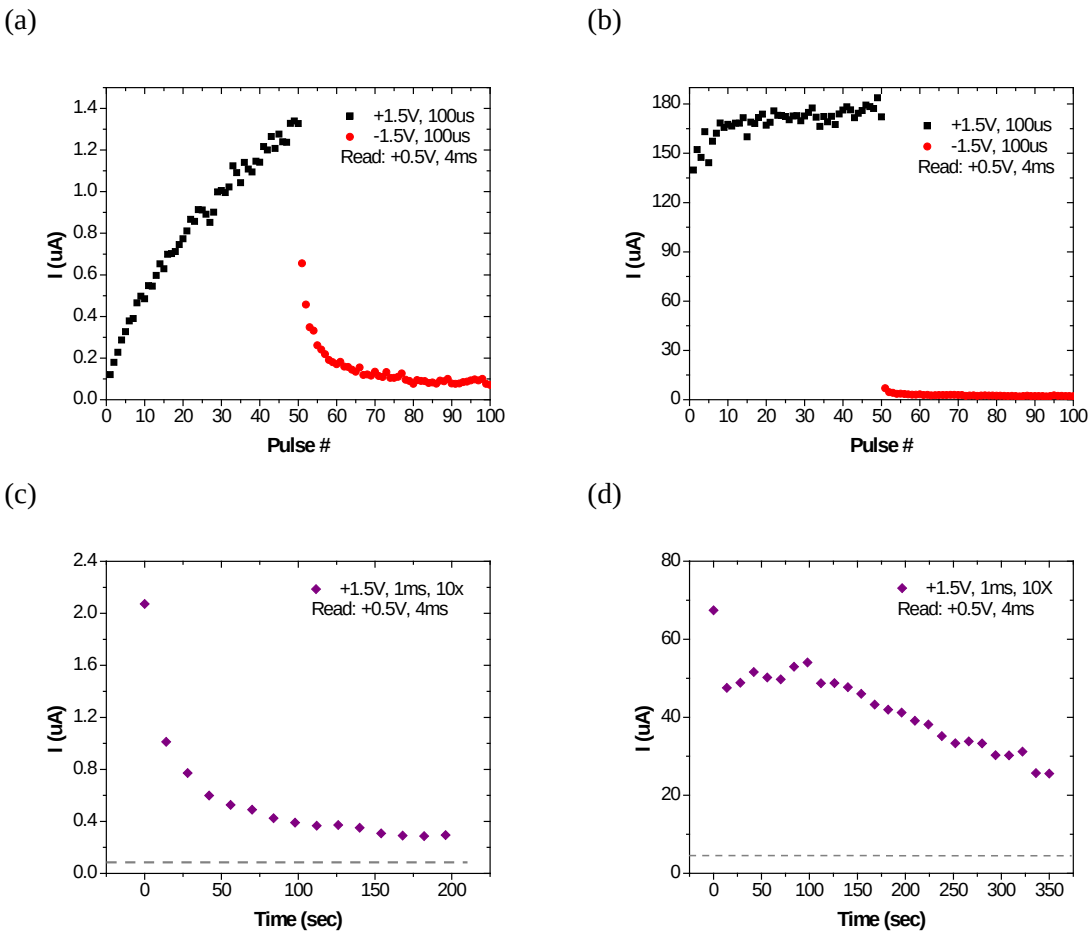


Figure 2.8 Analog switching vs. digital switching. By applying strong positive signals at higher CC, switching can be transitioned from (a) analog to (b) digital. Accompanied by the analog-to-digital switching is the elongated retention, as shown in (c) and (d), respectively. The retention curves are obtained by first potentiating the device with the same pulsing condition and then reading the decaying conductance periodically. Gray lines indicate the current level before potentiating pulses. Device size: 350 nm x 350 nm.

2.4.5 Current Rectification

In many samples, current rectification between bias polarities has been observed. Figure 2.9 shows one such example where the negative current is much smaller than the positive current. This current rectification is extremely favorable for memristor arrays because it eliminates the sneak path problem [27-28]. However, more systematic studies are still required to understand the rectification behavior.

(a) (b)

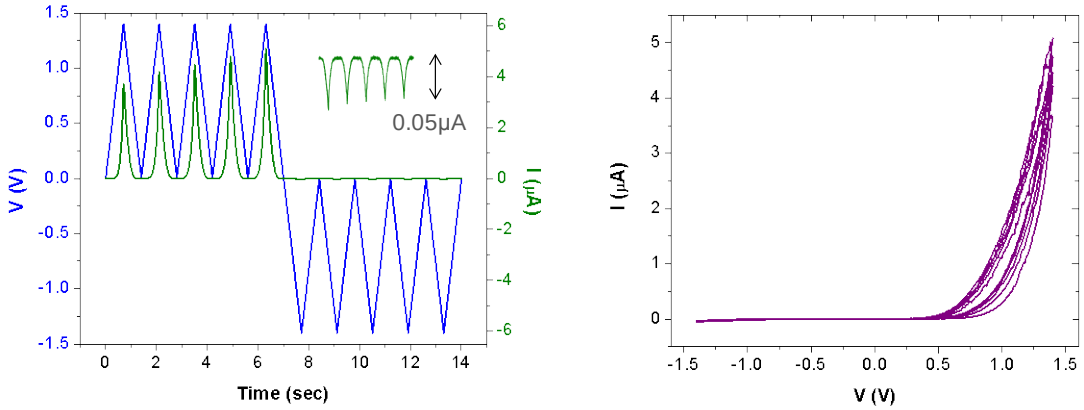


Figure 2.9 Current rectification in a WO_X memristor. (a) 5 consecutive positive followed by 5 consecutive negative DC sweeps. Inset shows the magnified negative current. (b) 5 alternating positive-negative DC sweeps.

2.5 Conclusion

The motivations and advantages of using WO_X as the memristor switching material were introduced. Briefly, compatibility of W with CMOS processing and low-temperature formation/deposition of WO_X make this material system readily integrate-able with the existing transistor technology. The simple crossbar device structure is easy to fabricate with high yields. Most importantly, switching is analog, bipolar, and stable; continuous modulation of device conductance can be well-controlled at reasonable voltage supply and current level, which makes it an ideal candidate for implementing important synaptic functions (e.g. potentiation and depression) at large scale. Moreover, WO_X memristors can endure over millions of cycles but with limited retention. Reasonable device-to-device and operation-to-operation variations are demonstrated even with pulse operations. Digital switching is also supported in these devices which makes

them useful for pure memory applications. Current rectification has been observed in some devices; although useful in memristor arrays, the mechanism behind the current rectification still needs to be systematically characterized and understood.

Reference

[1] J. D. Plummer, M. D. Deal, P. B. Griffin, Silicon *VLSI Technology: Fundamentals, Practice, and Modeling*; Prentice Hall, 2000.

[2] R. Sohal, C. Walczyk, P Zaumseil, D. Wolansky, A. Fox, B. Tillack, H. J. Müssig, T. Schroeder, "Thermal Oxidation of Chemical Vapour Deposited Tungsten Layers on Silicon Substrates for Embedded Non-Volatile Memory Application," *Thin Solid Films* **517**, 4534, 2009.

[3] W. C. Chien, Y. C. Chen, E. K. Lai, F. M. Lee, Y. Y. Lin, A. T. H. Chuang, K. P. Chang, Y. D. Yao, T. H. Chou, H. M. Lin, M. H. Lee, Y. H. Shih, K. Y. Hsieh, C. Y. Lu, "A Study of the Switching Mechanism and Electrode Material of Fully CMOS Compatible Tungsten Oxide ReRAM," *Appl. Phys. A –Mater.* **102**, 901, 2011.

[4] K. P. Biju, X. Liu, M. Siddik, S. Kim, J. Shin, I. Kim, A. Ignatiev, H. Huang, "Resistive Switching Characteristics and Mechanism of Thermally Grown WO_X Thin Films," *J. Appl. Phys.* **110**, 064505, 2011.

[5] C. H. Ho, C. L. Hsu, C. C. Chen, J. T. Liu, C. S. Wu, C. C. Huang, C. Hu, F. L. Yang, "9 nm Half-Pitch Functional Resistive Memory Cell with < 1 μA Programming Current Using Thermally Oxidized Sub-Stoichiometric WO_X Film," *IEDM*, San Francisco, CA, December 2, 2010.

[6] J. L. He, M. C. Chiu, "Effect of Oxygen on the Electrochromism of RF Reactive Magnetron Sputter Deposited Tungsten Oxide," *Surf. Coat. Tech.* **127**, 43, 2000.

[7] C. Bittencourt, R. Landers, E. Llobet, X. Correig, J. Calderer, "The Role of Oxygen Partial Pressure and Annealing Temperature on the Formation of W=O Bonds in Thin WO_3 Films," *Semicond. Sci. Technol.* **17**, 522, 2002.

[8] R. Yang, K. Terabe, T. Tsuruoka, T. Hasegawa, M. Aono, "Oxygen Migration Process in the Interface during Bipolar Resistance Switching Behavior of WO_{3-x}-Based Nanoionics Devices," *Appl. Phys. Lett.* **100**, 231603, 2012.

[9] K. Miyake, H. Kaneko, M. Sano, N. Suedomi, "Physical and Electrochromic Properties of the Amorphous and Crystalline Tungsten Oxide Thick Films Prepared under Reducing Atmosphere," *J. Appl. Phys.* **55**, 2747, 1984.

[10] C. E. Tracy, D. K. Benson, "Preparation of Amorphous Electrochromic Tungsten Oxide and Molybdenum Oxide by Plasma Enhanced Chemical Vapor Deposition," *J. Vac. Sci. Technol. A* **4**, 2377, 1986.

[11] D. S. Shang, L. Shi, J. R. Sun, B. G. Shen, F. Zhuge, R. W. Li, Y. G. Zhao, "Improvement of Reproducible Resistance Switching in Polycrystalline Tungsten Oxide Films by *in situ* Oxygen Annealing," *Appl. Phys. Lett.* **96**, 072103, 2010.

[12] N. Özer, "Optical and Electrochemical Characteristics of Sol-Gel Deposited Tungsten Oxide Films: A Comparison," *Thin Solid Films* **304**, 310, 1997.

[13] F. E. Rizzo, L. R. Bidwell, D. F. Frank, "Thermodynamics of the Tungsten-Oxygen System," *Matellurg. Soc. AIME Trans.* **239**, 1901, 1967.

[14] H. A. Wriedt, "The O-W (Oxygen-Tungsten) System," *J. Phase. Equil.* **10**, 368, 1989.

[15] S. Santucci, C. Cantalini, M. Crivellari, L. Lozzi, L. Ottaviano, M. Passacantando, "X-Ray Photoemission Spectroscopy and Scanning Tunneling Spectroscopy Study on the Thermal Stability of WO_3 Thin Films," *J. Vac. Sci. Technol. A* **18**, 1077, 2000.

[16] J. Meyer, S. Hamwi, M. Kröger, W. Kowalsky, T. Riedl, A. Kahn, "Transition Metal Oxides for Organic Electronics: Energetics, Device Physics, and Applications," *Adv. Mater.*, DOI: 10.1002/adma.201201630.

[17] D. B. Migas, V. L. Shaposhnikov, V. E. Borisenko, "Tungsten Oxides. II. The Metallic Nature of Magnéli Phases," *J. Appl. Phys.* **108**, 093714, 2010.

[18] T. Chang, S. H. Jo, K. H. Kim, P. Sheridan, S. Gaba, W. Lu, "Synaptic Behaviors and Modeling of A Metal Oxide Memristive Device," *Appl. Phys. A – Mater.* **102**, 857, 2011.

[19] W. M. Cowan, Ed. *Synapses*; Johns Hopkins University Press, 2001.

[20] ITRS Roadmap, 2011 Edition.

[21] A. C. Torrezan, J. P. Strachan, G. Medeiros-Ribeiro, R. S. Williams, "Sub-Nanosecond Switching of A Tantalum Oxide Memristor," *Nanotechnology* **22**, 485203, 2011.

[22] P. Sheridan, *unpublished*.

[23] J. J. Yang, M. X. Zhang, J. P. Strachan, F. Miao, M. D. Pickett, R. D. Kelley, G. Medeiros-Ribeiro, R. S. Williams, "High Switching Endurance in TaO_X Memristive Devices," *Appl. Phys. Lett.* **97**, 232102, 2010.

[24] F. Miao, J. J. Yang, J. Borghetti, G. Medeiros-Ribeiro, R. S. Williams, "Observation of Two Resistance Switching Modes in TiO_2 Memristive Devices Electroformed at Low Current," *Nanotechnology* **22**, 254007, 2011.

[25] J. Hermiz, T. Chang, C. Du, *in preparation*.

[26] E. Bullmore, O. Sporns, "Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems," *Nat. Rev. Neurosci.* **10**, 186, 2009.

[27] K. H. Kim, S. H. Jo, S. Gaba, W. Lu, “Nanoscale Resistive Memory with Intrinsic Diode Characteristics and Long Endurance,” *Appl. Phys. Lett.* **96**, 053106, 2010.

[28] S. C. Puthentheradam, D. K. Schroder, M. N. Kozicki, “Inherent Diode Isolation in Programmable Metallization Cell Resistive Memory Elements,” *Appl. Phys. A. – Mater.* **102**, 817, 2011.

Chapter 3

Material and Electrical Studies of WO_X Memristors

3.1 Introduction

The earliest reports on RS in oxide insulators dated back to the 1960s [1]. Throughout the decades, more and more materials have been found to show RS in a simple MIM structure, including transition metal oxides (TMOs), chalcogenides, electrolytes, glass, semiconductors, and polymers. A categorization of RS effects is summarized in Figure 3.1 [2, 3].

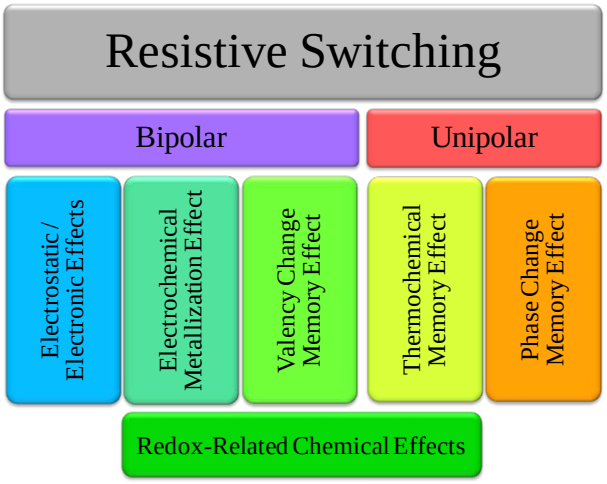


Figure 3.1 Resistive switching effects classified according to operation modes and switching mechanisms. Modified from [3].

In a very straightforward manner, RS effects can be divided into *bipolar* and *unipolar* modes. In the bipolar mode, the SET voltage to turn ON the memory device and the RESET voltage to turn OFF the memory device are of opposite polarities, whereas in the unipolar mode, SET and RESET voltages are of the same polarity with the ON/OFF switching controlled by the external current compliance (CC). The above explanation holds for digital as well as for analog memory devices. To be more specific, RS effects can furthermore be classified according to their physical mechanisms. Listed here are five mechanisms that have received most attention. A brief description of each effect is provided as the following [3]:

- Electrostatic / Electronic Effects: effects involving only electrons, including charge-trapping, Fowler-Nordheim (F-N) tunneling, interface states trapping, thermionic emission, and metal-insulation-transition (MIT).
- Electrochemical Metallization Effect (ECM): switching achieved through the reduction-oxidation (redox) reaction of the active electrode material and transportation of metal cations across the solid-electrolyte film; also known as Conductive Bridging RAM (CBRAM) or Programmable Metallization Cell (PMC).
- Valency-Change Memory Effect (VCM): anion migration (e.g. V_{OX}) that changes the valency and thus the stoichiometry of TMOs. Here similar redox processes are involved, but the participating species (e.g. V_{OX}) already exist in the switching film and active electrodes are not required.
- Thermochemical Memory Effect (TCM): stoichiometry change induced by Joule heating in TMOs.

- Phase Change Memory Effect (PCM): crystallinity change controlled by quenching speed in chacolgenide materials.

Caution should be exercised because in any system, multiple mechanisms might co-exist and exchange dominant roles depending on operations. In particular, ECM, VCM, and TCM effects all rely on redox-related electrochemical effects. Hence one should be very careful when attempting to identify the switching mechanisms and when interpreting experimental results.

3.2 Material Study

3.2.1 Stoichiometry and Energy Bands

Similar to most other TMOs, many WO_X compounds have been recognized. Common stable oxide states are WO_3, $WO_{2.9}$ (in the form of $W_{20}O_{58}$), $WO_{2.72}$ (in the form of $W_{18}O_{49}$), and WO_2 [4-5]. In addition to stable states, there are metastable states referred to as the "Magnéli phases" which comprise a series of substoichiometric phases. From the view of electronic configurations, substoichiometric phases are the reduced forms of oxides, that is, there is oxygen deficiency in the materials. WO_3 (i.e. the stoichiometric form of tungsten oxide) is a typical d^0 TMO whose valence band maxima are composed of O 2*p* states and conduction band minima of W 5*d* states [6]. The *d*-band is empty in the case of fully oxidized cations (W^{6+}) but tends to be partially filled by electrons due to the presence of oxygen vacancy defects. This also results in donor states within the bandgap that shift the Fermi level closer to the conduction band, making WO_X

n-type semiconductors (Figure 3.2). As a consequence, the bandgap, work function, and crystal structure are altogether affected.

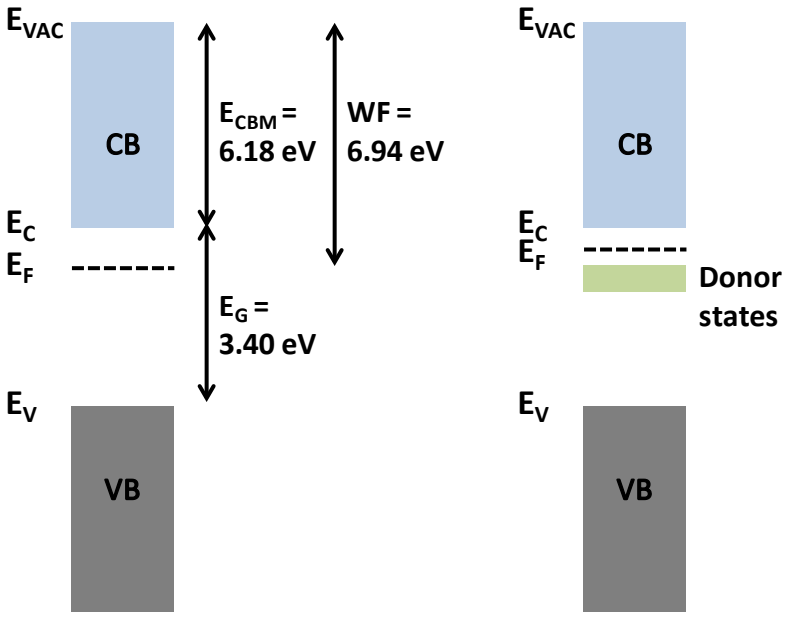


Figure 3.2 Band diagrams of (left) stoichiometric WO_3 and (right) oxygen deficient WO_X. E_{VAC}, E_F, E_{CBM}, E_G, WF denote the vacuum level, Fermi level, conduction band minimum energy, bandgap, and work function, respectively. Donor states push E_F closer to E_C. Modified from [6].

The bandgap of WO_3 is ~ 3.4 eV but can be as low as 2.6 eV when reduced. The work function (WF) of WO_3 is ~ 6.9 eV which also drops towards that of W (~ 4.6 eV) as the material gets less stoichiometric. Moreover, interfacial effects and surface adsorption are additional factors that can alter the band structures [6-7]. For example, ideally there should be no Schottky barriers formed at the Pd/WO_X and W/WO_X interfaces because WO_X has a higher WF than those of Pd (~ 5.6 eV) and W. However, it is argued that oxygen deficiency, interfacial states, and moisture adsorption all act to reduce the WF of WO_X such that there may actually be Schottky barriers formed at both

interfaces [6-9]. This hypothesis is supported by the nonlinear DC *I-V* curves in Figure 2.3 where minimal current flows below ±0.2 V.

3.2.2 Structural Analysis through X-Ray Diffraction (XRD)

XRD was employed to investigate the crystallinity of WO_X films prepared under different conditions. Tungsten oxide films are usually amorphous when prepared below 300 °C but begin to crystallize at > 350 °C [10]. In addition, for thermally-oxidized WO_X, the higher the oxidation temperature and the longer the oxidation time, the more crystallized the film becomes. In polycrystalline WO_X films, it is believed that grain boundaries (GBs) are the locations where V_{OX} segregate to and transport faster, thus it is useful to study the crystallinity of the films that show RS behaviors [11].

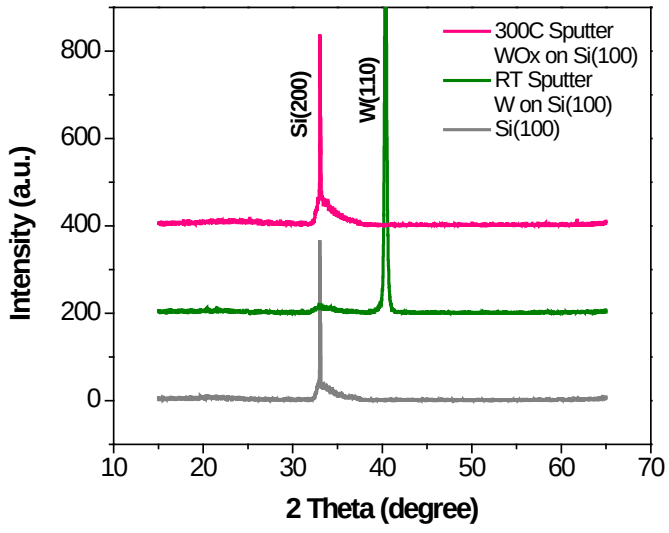


Figure 3.3 XRD patterns of Si substrate (bottom trace), W film (middle trace), and sputtered WO_X film (top trace). The traces are offset for clarity. The reference data used are based on monoclinic WO_3 measurements.

Figure 3.3 shows the XRD patterns of the substrate, sputtered W, and sputtered WO_X. The Si(100) substrate ideally should not render any peak in the degree range but the Si(200) peak near 34° is very often seen due to imperfect sample-mounting [10]. The sputtered W film has a typical peak near 42° suggesting a polycrystalline (110) structure. The sputtered WO_X film was prepared by dc sputtering from a W metal target in an Ar/O_2 gas mixture (in this sample, 30 % O_2) at 300 °C. Because 300 °C is not sufficient for crystallization, the sputtered WO_X is amorphous, displaying only the Si(200) peak.

However, at temperatures above 350 °C WO_X begins to crystallize. Figure 3.4 shows the XRD patterns of RTA-oxidized WO_X films of which WO_3 peaks near 23° and 33° are seen. For both peaks, multiple orientations may be superimposed and unresolved; for example, the WO_3 (002), (020), and (200) all show peaks near 23° [10]. WO_X oxidized at 450 °C is more crystallized than that at 350 °C; also, WO_X oxidized for 3 min is more crystallized than that for 1 min. Both observations agree with the tendency reported in [10]. Although not very clear, a $WO_{2.92}$ peak may be identified in addition to the WO_3 peaks in Figure 3.4(c)-(d). As a result, we expect that RTA not only oxidizes W into WO_X but at the same time crystallizes WO_X to form larger grains. As the WOx film grows from the WOx/W interface, the film should become more crystallized near the top surface because that part has been subjected to oxidation longer. This information is important to understanding the memristor operations as V_{OX} segregate to GBs and transport much faster along GBs [10-11].

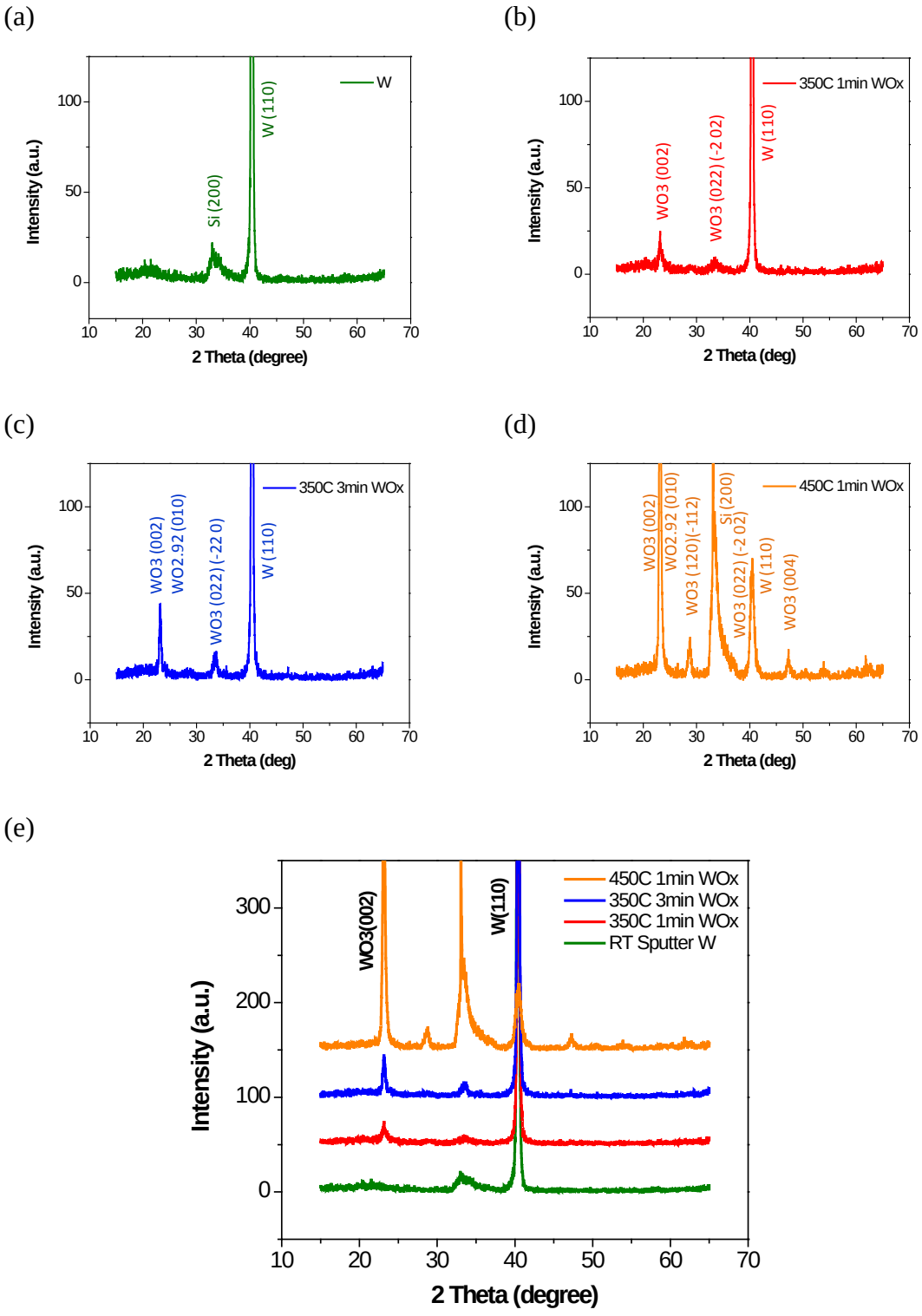


Figure 3.4 XRD patterns of W and RTA-oxidized WO_X films prepared at different oxidation temperature and time. (a) and (d) may have signals from Si(200) superimposed. Traces in (e) are offset for clarity.

3.2.3 Chemical Analysis through X-Ray Photoelectron Spectroscopy (XPS)

XPS is a common surface chemical technique that yields quantitative information on elemental composition and valence states of the sample through binding energy characterization. Figure 3.5 shows the W4*f* core spectra where the two peaks are due to the $4f_{7/2}$ and doublets. According to literature [9-10, 12-13], the position of the doublets for stoichiometric WO_3 are at 35.8 eV ($4f_{7/2}$) and 37.9 eV ($4f_{5/2}$), corresponding to the W^{6+} oxidation state; while the position of the doublets for W metal are at 31.3 eV ($4f_{7/2}$) and 33.4 eV ($4f_{5/2}$), corresponding to the W^0 oxidation state. The spin-orbit-splitting is about 2.14 eV and the energy shift between WO_3 and W is about 4.5 eV. Between the W^{6+} and W^0 doublets, tungsten may also appear in W^{5+}, W^{4+}, and W^{2+} oxidation states. Primitive XPS scans suggest that the RTA-oxidized WO_X film surface has a composition close to WO_3 (i.e. near stoichiometric). Here, contributions from W^{5+} and W^{4+} may be too small to be resolved. In general, one should not expect to be able to accurately quantify substoichiometry merely through peak-fitting since there are too many parameters which affect the fitting process. Meanwhile, sputtered WO_X film showed obvious evolving peaks of W^{5+} and W^{4+} when the O_2 percentage of the sputtering gas mixture was decreased from 35 % to 15 % [14]; the results are provided in the Appendix at the end of this chapter for convenience.

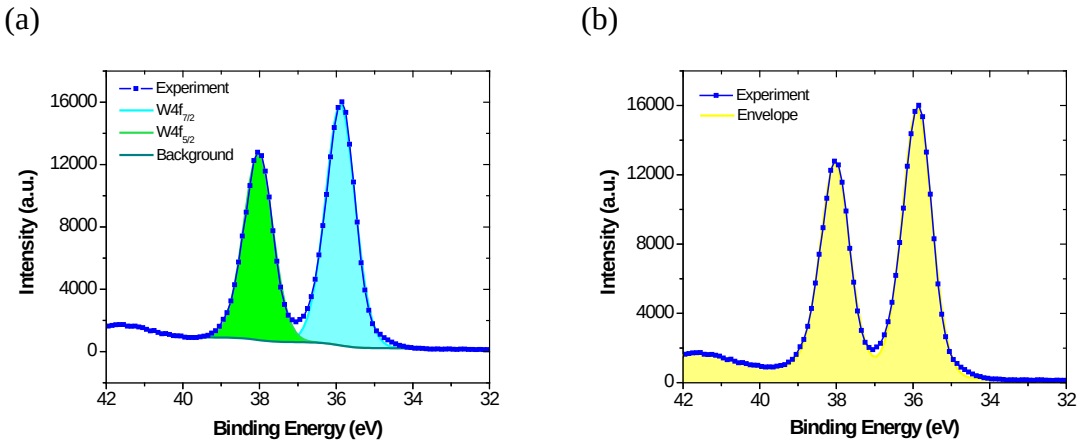


Figure 3.5 XPS W4*f* core spectra. (a) Experimental data and fitting curves. The peaks are fitted using Gaussian-Lorentzian curves and Shirley type background. (b) Experimental data and the final fitting envelope. The WO_X film used for XPS study was RTA-oxidized at 400 °C for 2 min. Note that the RTA tool has been fitted with different types of thermalcoupler and has been calibrated infrequently, so this condition actually renders WO_X films similar to the 350 °C samples oxidized at other times.

3.3 Electrical Study

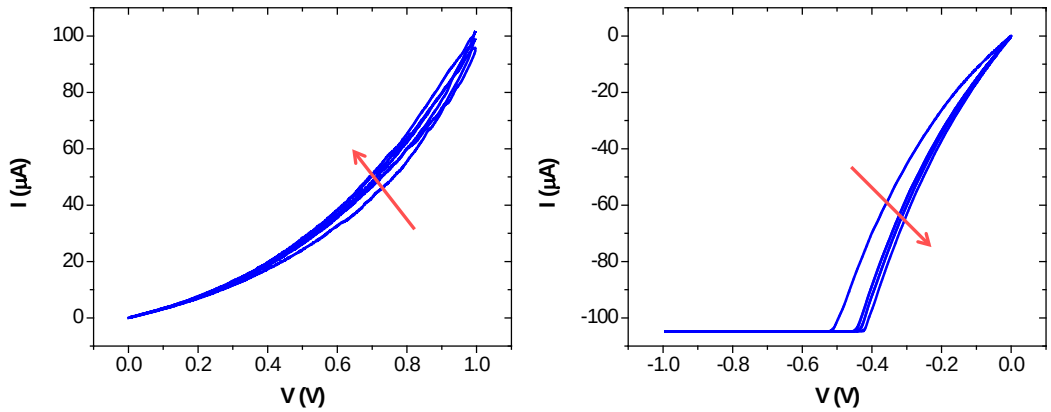
3.3.1 Effects of Annealing Gases

In order to study the effects of annealing under different gas compositions, a sputtered WO_X film was used as the switching layer and was assumed to be stoichiometrically uniform throughout its thickness to first order. Both electrodes were inert (Pd). Surface annealing treatments in N_2, Ar, and O_2 were given to change the film stoichiometry and to create a V_{OX} gradient along the film thickness. Fabrication steps for this device structure are provided in the Appendix at the end of this chapter.

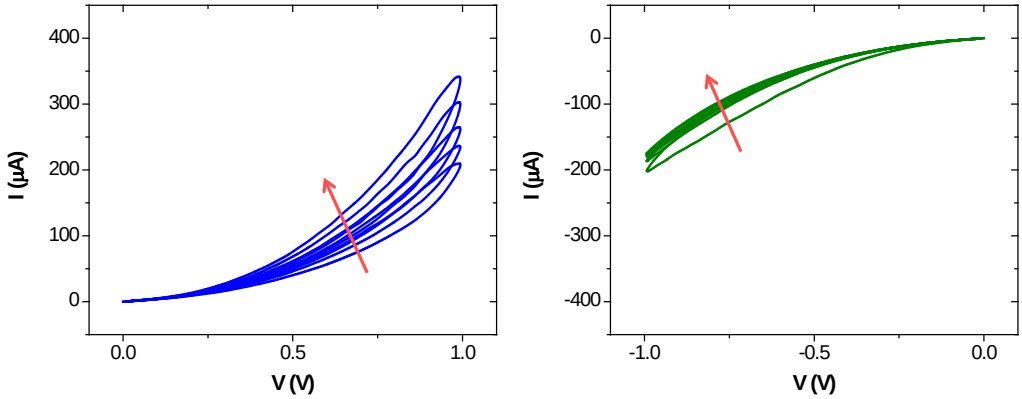
The non-treated samples had a symmetric Pd/WO_X/Pd structure with a uniformly sputtered WO_X film so switching was unstable with indefinite polarity (Figure 3.6(a)). Annealing in inert gases such as N_2 and Ar significantly changed the switching behaviors. It is uncertain whether these inert gases are capable of reducing WO_X, but oxygen could

also easily escape from the WO_X surface at elevated temperature due to diffusion. Thus, a great amount of V_{OX} was generated in the film, especially near the surface. The V_{OX} gradient determined the switching polarity; a positive bias drove V_{OX} towards BE gradually increasing the conductance, while a negative bias attracted them towards the TE gradually decreasing the conductance (Figure 3.6(b)-(c)). For brevity, this switching polarity is noted as "+P/-D" where P and D denote potentiation and depression, respectively. Also, the current increased by an order of magnitude compared to the untreated samples, confirming the high concentration of V_{OX} introduced by N_2 or Ar annealing. In contrast, annealing in oxidizing ambient resulted in opposite effects by annihilating V_{OX} near the surface. As can be seen from Figure 3.6(d), the switching polarity was reversed (-P/+D) and the current level dropped by two orders of magnitude. Based on the same argument, the switching polarity indicated a more stoichiometric, insulating top surface. Results from this study are consistent with the V_{OX} model and explained how annealing creates/annihilates V_{OX} near the film surface and how the V_{OX} distribution affects the switching polarity.

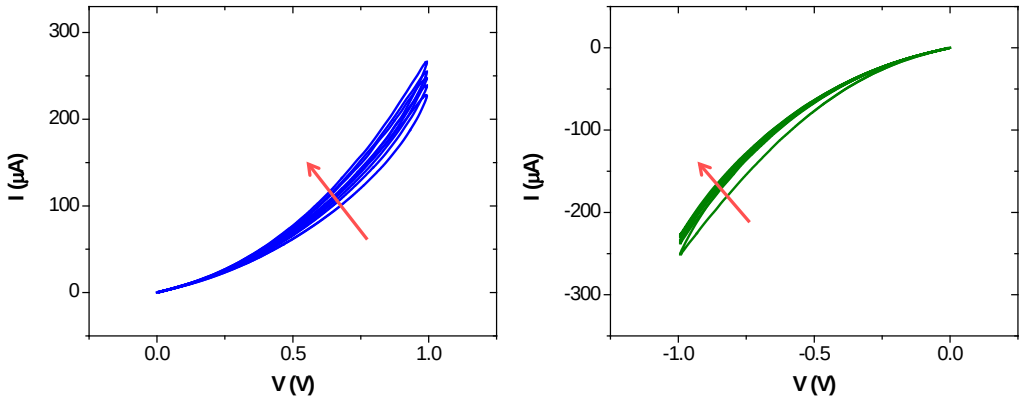
(a) Non-treated



(b) N_2, 400 °C, 1 min



(c) Ar, 400 °C, 1 min



(d) O_2, 400 °C, 1 min

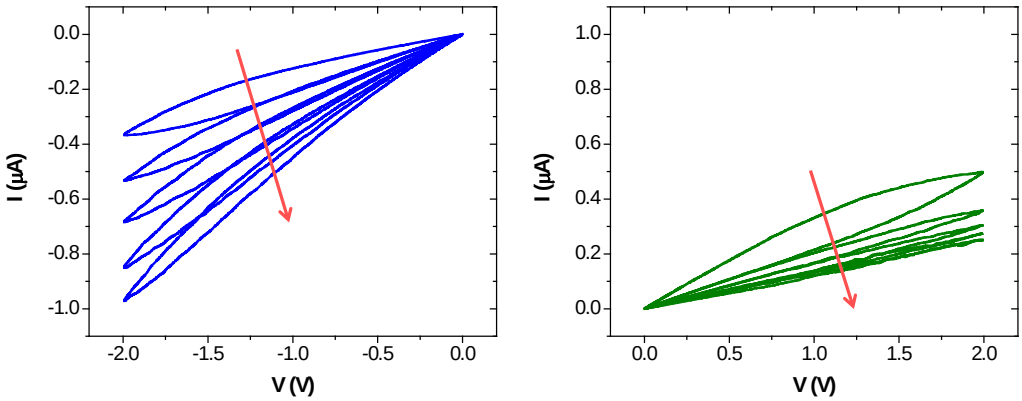


Figure 3.6 Effects of annealing gases. (a) No annealing, (b) N_2 annealing, (c) Ar annealing, and (d) O_2 annealing. Switching polarity and current level were strongly affected. Arrows indicate switching direction. Device sizes: 200 nm x 200nm in (a)-(c), and 150 nm x 150 nm in (d).

3.3.2 Effects of Electrode Materials

It is known that different metals react differently when in contact with semiconductor or dielectric materials. To understand the effects of TE materials on the TE/WO_X interface, an Ellingham diagram is useful in providing information on whether a reaction is thermodynamically favorable [15-16]. In Figure 3.7, any lines that are below the W-O (WO_2 and WO_3) lines are naturally favorable reactions (e.g. Al, Hf, Ti, Ta, Cr) and any lines that lie above the W-O lines are unfavorable reactions and thus should not occur spontaneously (e.g. Ag, Pt, Pd). Studies on TiO_X/metal interfaces have given much support to this hypothesis [16-17]. As mentioned earlier, due to the presence of interface states, surface adsorption, and potential redox reactions between the electrode and WO_X, the WO_X/metal interface characteristics could not be correctly predicted by merely judging from the metal WFs. Rather, free energy of formation, expressing the energy cost of a metal material being oxidized, is a more suitable index to describe the device behavior. Besides free energy of formation, oxidation potential, oxygen affinity, and electronegativity are also used to refer to the same concept [7]. In Figure 3.8, metals with low energies of formation generally show Ohmic conduction at the metal-oxide/metal junction while metals with high energies of formation in principle show Schottky-like rectifying conduction [17].

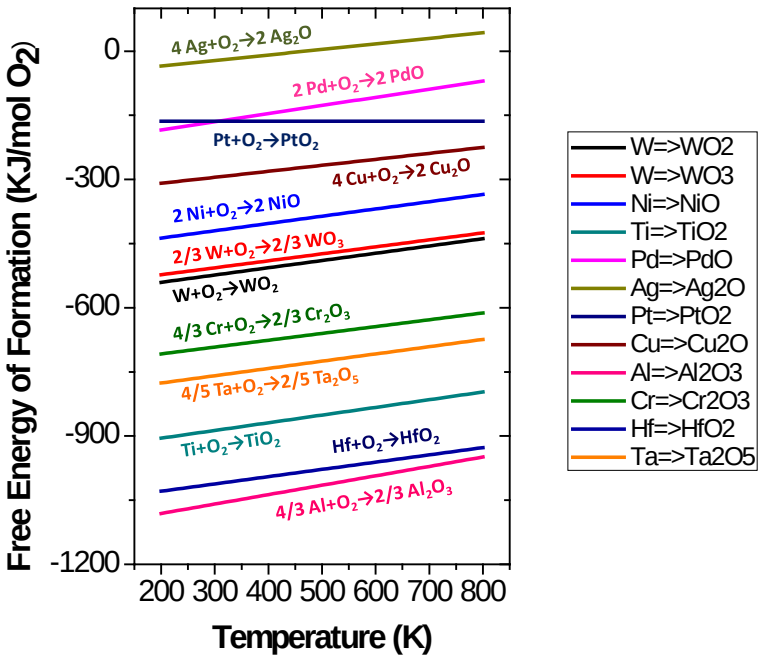


Figure 3.7 The Ellingham diagram for various WO_X sub-oxide reactions and other metal oxides. Equations and constants for materials were from [18] and [19]. Note that the energy for Au is positive and is not included in this figure.

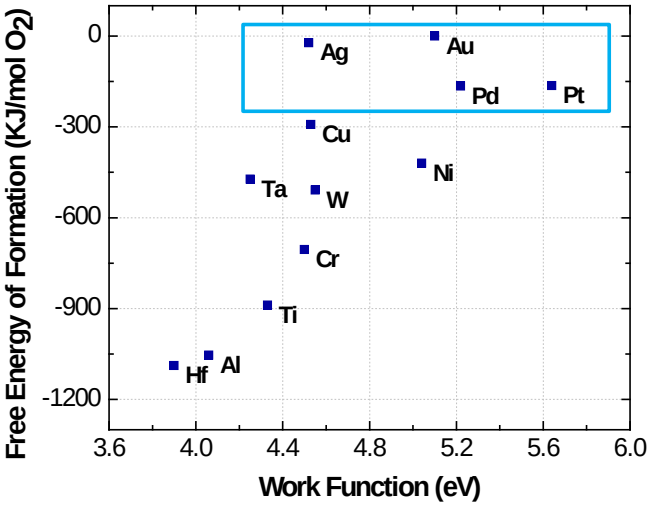
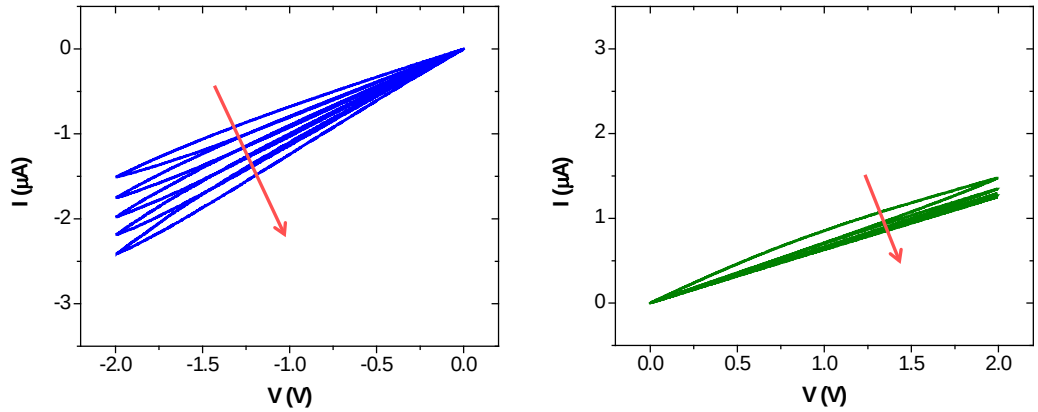


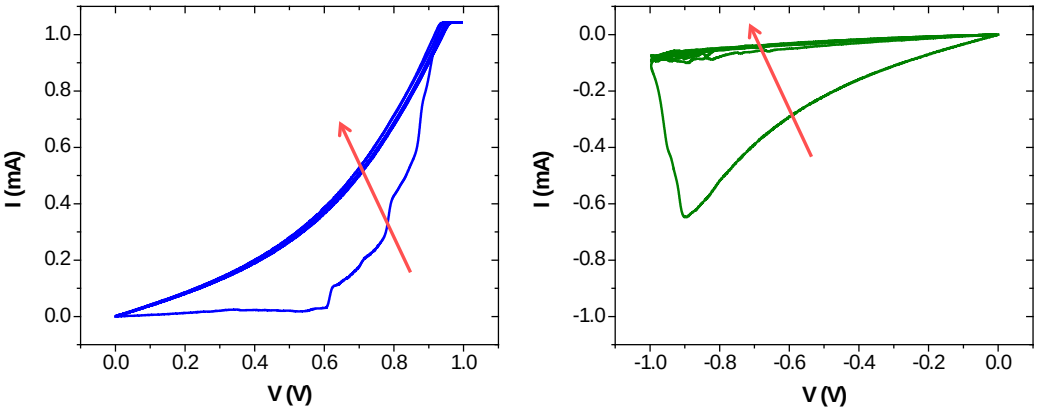
Figure 3.8 Free energies of formation and work functions of common metals. Metals in the bracket show rectifying conduction when in contact with WO_X. Constants were from [18].

In this study, Au, Ni, Al, and Ti were used as TE materials on O_2-annealed WO_X. Fabrication steps are given in the Appendix at the end of this chapter. According to the Ellingham diagram, Au is more stable than W as Pd and Pt are, hence should not induce WO_X reduction near the WO_X/TE interface. Figure 3.9(a) shows switching of the Au samples which resembled the Pd samples (Figure 3.6(d)). In contrast, Ni, Al, and Ti have similar or lower energies of formation than that of W, hence should spontaneously reduce WO_X at the WO_X/TE interface. Figure 3.9(b)-(c) agree with the speculation that both Ni and Al act as reducing agents creating V_{OX} near the surface, so the switching polarity was the same as those in Figure 3.6(b)-(c), where the WO_X surface had undergone an inert gas annealing. Compared to the Ni samples, the Al samples showed higher conductance at fresh and could not be switched back to the original state once brought to higher conducting states. Due to the high current level (~ 1 mA) analog switching was no longer observed in Ni and Al samples. The Ti samples should render similar switching behavior, and indeed, all devices showed no switching but shorted regardless of their sizes. Interestingly, the WO_X/metal reaction is more effective than annealing because the Ni and Al samples overcame the switching polarity caused by O_2 annealing. The results from annealing gas and TE material studies are summarized in Table 3.1.

(a) Au TE



(b) Ni TE



(c) Al TE

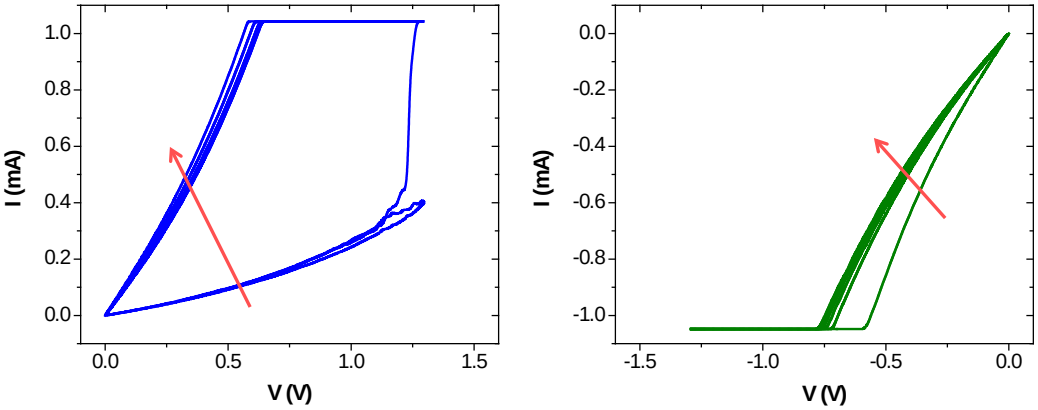


Figure 3.9 Effects of TE materials. (a) Au, (b) Ni, and (c) Al. Switching polarity and current level were strongly affected. Arrows indicate switching direction. Device size: 150 nm x 150 nm in (a)-(b), and 200 nm x 200 nm in (c).

Table 3.1 Summary of the effects of annealing gases and electrode materials.

Annealing	**TE**	**P / D**	**I**	**Comments**
N/A	Pd	unstable	~ 100 uA	Symmetrical structure
N_2	Pd	+ / -	~ 100 uA	Create V_{OX} near TE Interface
Ar	Pd	+ / -	~ 100 uA	Create V_{OX} near TE interface
O_2	Pd	- / +	~ 1 uA	Annihilate V_{OX} near TE interface
O_2	Au	- /+	~ 1 uA	Annihilate V_{OX} near TE interface
O_2	Ni	+ / -	~ 1 mA	Create V_{OX} near TE interface
O_2	Al	+ / -	~ 1 mA	Create V_{OX} near TE interface
O_2	Ti	+ / -	short	Create V_{OX} near TE interface

3.3.3 Effects of Testing Environments

Since RS in TMOs is strongly related to the oxidation states, the concentration and distribution of V_{OX} within the switching material are critical. There have been studies reporting extremely distinct switching behaviors of TMOs under vacuum and air [20-21]. In particular, it has been argued that oxygen in air and adsorbed in the Pt electrode plays a key role in the switching of the WO_X thin film supported by the fact that no stable bipolar switching was observed under vacuum or when Pt is replaced with Au [20]. It is well-accepted that the WO_X film quality can vary drastically when different preparation methods are used [5, 22], plus the metal electrode materials and interface effects that are present in the device structure, so far no explanations in the literature can be directly applied to the WO_X devices developed here.

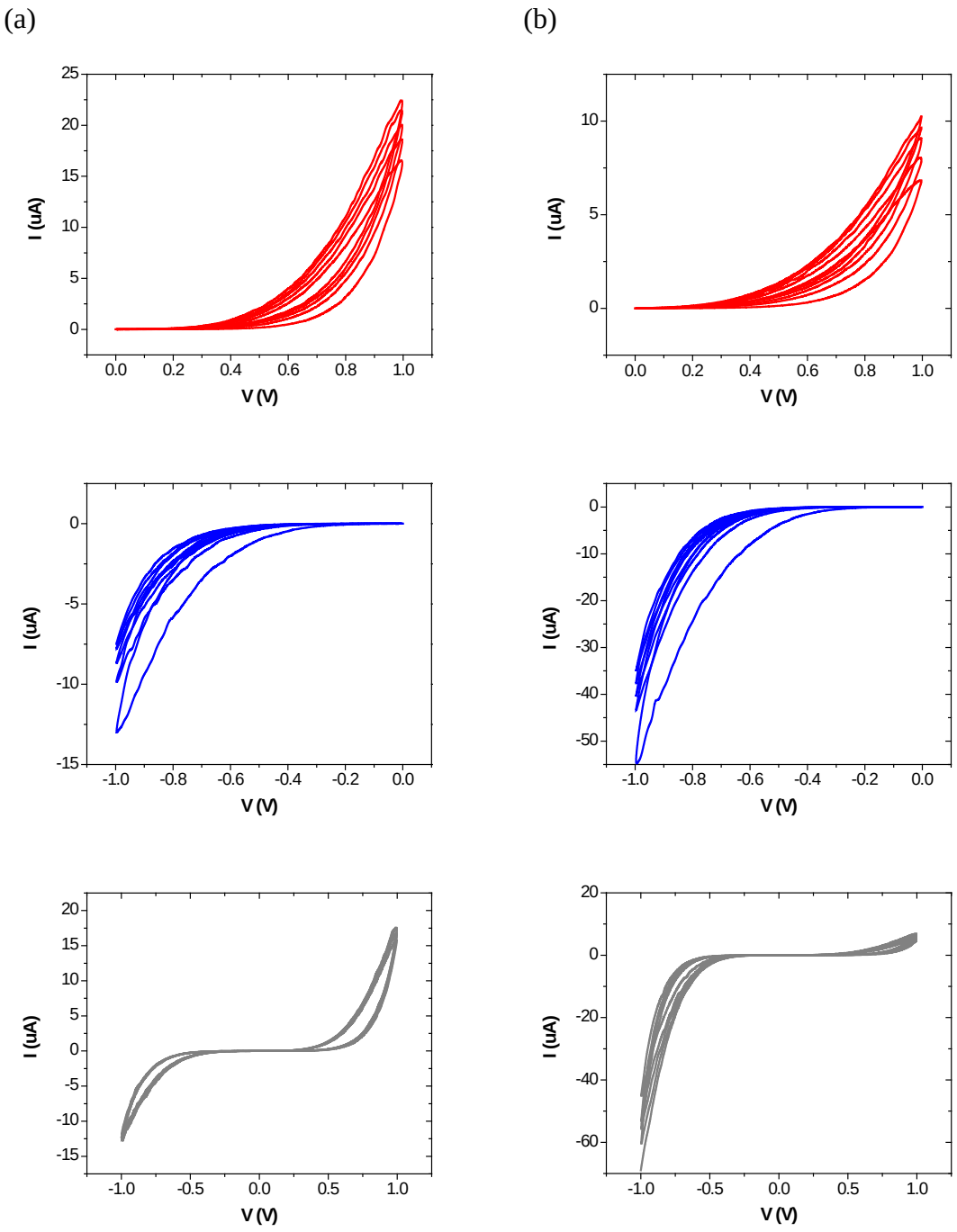


Figure 3.10 DC *I-V* switching curves in (a) air and (b) vacuum (~0.4 mTorr). Red: 0 V → 1 V → 0 V, 5 cycles; blue: 0 V → -1 V → 0 V, 5 cycles; Gray: -1V → 1 V → -1 V, 5 cycles. Sweep rate was 2 V/s for all tests.

Figure 3.10(a) and (b) show the typical DC *I-V* curves when measured in air and in vacuum, respectively. Stable analog switching is seen in both cases but with different

current rectification. These results may suggest two different factors that take place upon the deprivation of oxygen. First, larger currents in the negative polarity (blue curves) were observed and can be explained by the fact that the V_{OX} concentration in the film was increased and thus the overall film conductivity was increased. Second, smaller currents in the positive polarity (red curves) were observed even though the film was overall more conductive; this effect can be explained by the fact that the work function of Pd was increased in vacuum and thus the Schottky barrier height at the TE/WO_X interfaces increased [6]. The rectification behavior can be more easily seen with the gray curves.

3.4 Effect of Joule Heating

It has been known that in TMO RRAM devices current-induced Joule heating leads to positive feedback effects during RS that either speed up V_{OX} transport or cause structural changes leading to enhanced conduction [23-26]. In the particular case of tungsten oxides, heating may cause both effects. According to the phase diagram [4, 27-28], tungsten oxide is a polymorphic, meaning that different phases of tungsten oxides can co-exist at a fixed temperature; for instance, WO_2 and WO_3 co-exist below 484 °C, WO_2 and $WO_{2.9}$ co-exist between 484 °C and 585 °C, and WO_2 and $WO_{2.72}$ co-exist between 585 °C and 1530 °C. Even for a fixed composition ratio, multiple bonding structures have been observed.

To estimate the local temperature of the conducting filaments (or conducting phases) of the WO_X film in an actual device, Equation 3.1 is used [29]

$$\frac{dT}{dt} = \alpha \frac{V^2(t)}{R} - \frac{1}{\tau_C}(T - T_0) \tag{3.1}$$

where α reflects the heating term, R is the electrical resistance, and τ_C is the characteristic cooling time which depends on the geometry and thermal conductivity of the material. At steady-state, Equation 3.1 can be derived to calculate the local temperature

$$T - T_0 = \tau_C \alpha \frac{V^2}{R} \tag{3.2}$$

in which $\tau_C \alpha$ can be expressed as [29]

$$\tau_C \alpha = \frac{\ln(d/r)}{4 \cdot 18(2\pi KL)} \tag{3.3}$$

where r is the radius of each filament cylinder, d is the mean filament separation, K is the thermal conductivity, and L is the filament length. d is defined by $d^2N=1$ where N is the number of filaments per area. The exact value of (d/r) is not well-defined but does not make a huge difference because it appears as the argument of a log function. If K is taken to be 1.3 $Wm^{-1}K^{-1}$ [30] and L to be 10 nm, then

$$T - T_0 \approx 0.5 \cdot I \cdot V \tag{3.4}$$

with I in microampere and V in volt, and T_0 is taken to be the room temperature. For a typical analog operation (e.g. I ~ 20 μA and V ~ 1.2 V), temperature increases tens of degrees (K or °C); for digital operation (e.g. I ~ 200 μA and V ~ 1.5 V), temperature increases a few hundreds of degrees (K or °C).

Note that the above discussion is a rough thermodynamics approximation assuming a uniform/homogenous material, so it merely considers Joule heating and heat loss to the insulator excluding defect generation and/or phase transformation that

accompanied heating. Also, here only steady-state is considered, hence no information is inferred about the switching dynamics.

Another approach developed in [23] can also be used to estimate the local temperature which is derived assuming a metallic core/insulating shell structure conducting current in parallel. With appropriate boundary conditions, they arrived at

$$u = \exp[-\frac{1}{2}W_0(\frac{4\pi^2 K r_{chan}^2 (T - T_0)}{i^2 \rho_{met}})] \tag{3.5}$$

where r_{chan} is the radius of the outer edge of the shell, ρ_{met} is the resistivity of the metallic core, i is the current, and $W_0(x)$ is the zero-branch Lambert W function. u, the state variable, is defined as the normalized radius of the metallic core with respect to the insulating shell, which can be related to the density of conducting filaments. Assuming ρ_{met} = 3 x 10-6 Ωm, K = 1.3 $Wm^{-1}K^{-1}$, and r_{chan} = 100 nm, a relationship between u and (T-T_0) can be obtained numerically for both analog and digital operations, as shown in Figure 3.11. It is clearly seen that for analog switching, u is usually very small with a temperature increase on the order of tens of degrees, whereas for digital switching u is much larger with a temperature increase on the order of hundreds of degrees. Same as the previous approach, once again, this is a very rough approximation as K and ρ_{met} are actually temperature-dependent parameters, and the physical dimensions used for calculation are not experimentally determined; additionally, more uncertainty arises if phase change is considered. Nevertheless, one could draw a conservative conclusion that both methods gave similar trend and predicted the same orders of magnitude in temperature increase.

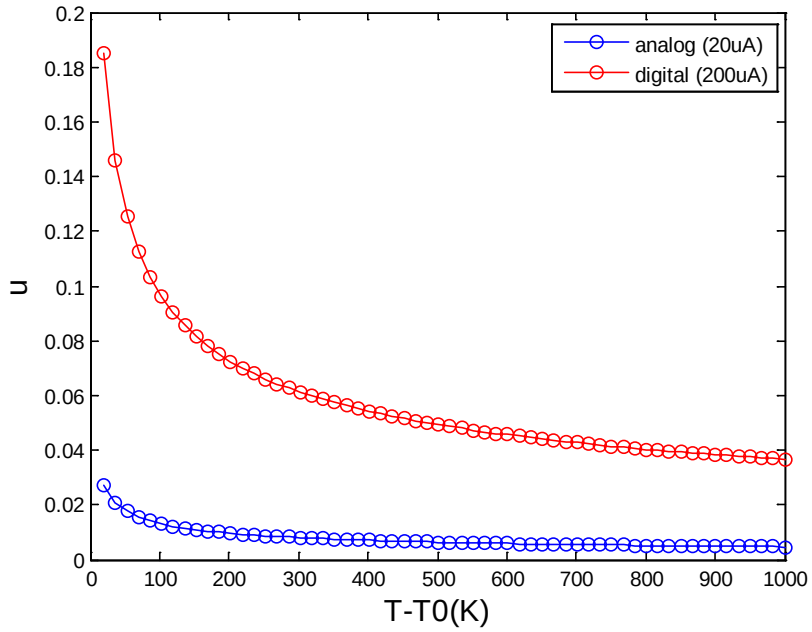


Figure 3.11 Heating of conducting metallic core as a function of filament density at digital and analog operating conditions.

3.5 Postulated Switching Mechanism in WO_X Memristors

Based on the understanding through designed experiments and established theory on conduction properties in various TMOs [2, 16, 23-26], we postulate here that the WO_X film, when fabricated into a memristor configuration using the RTA method, has a built-in asymmetry in which V_{OX} is more accumulated near the TE/WO_X interface. When positive voltage is applied, the electric field drives the positively-charged V_{OX} toward the WO_X/BE surface. Wherever the V_{OX} concentration is increased high enough, the local conductivities are significantly increased and large switching current is supported due to tunneling between the conducting paths and BE. Figure 3.12 illustrates this concept of a memristor in its resistive, intermediate, and conductive states. The V_{OX} transport can be further speeded up by the thermal effects during the switching process.

Additionally, as shown in Figure 3.12(b), natural diffusion of V_{OX} away from the conducting paths will degrade the device retention. On the other hand, as more/stronger positive bias is applied, more conductive paths are formed, as in Figure 3.12(c), and the conducting regions will less likely be destroyed by V_{OX} diffusion, leading to a better retention performance. This phenomenon will be discussed in Chapter 5 in detail.

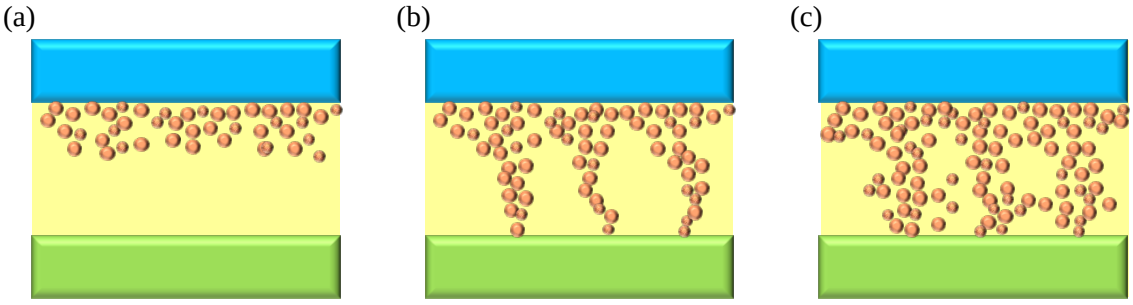


Figure 3.12 Illustrations of the different states in a WO_X memristor due to V_{OX} distribution. (a) Resistive state where the V_{OX} concentration is low near the WOx/BE interface, (b) intermediate state with a few conducting paths due to increased V_{OX} concentration in these regions, and (c) conductive states with many conducting paths (or enlarged areas of the conducting paths).

This switching polarity may seem counterintuitive because one would imagine the top surface to be more oxidized and the WO_X/BE to be partially oxidized. As discussed in the XRD analysis, the proposed V_{OX} gradient shown in Fig. 3.12 is due to the gradient in crystallinity during oxidation. Another possible explanation is that upon the application of positive bias on the TE, V_{OX} is generated near the TE/WO_X interface due to the escape of oxygen through the GBs of the Pd TE [9, 20].

3.6 Conclusion

Despite the overwhelming literature on tungsten oxides for various applications using diverse preparation methods, very little information can be directly applied to the films used in WO_X memristors due to the versatility of the material. XRD and XPS studies were carried out to understand the morphology and stoichiometry of the films, respectively. The RTA-oxidized films in the temperature range interested were identified to be polycrystalline and near-stoichiometric. Later, different annealing gases and metal contacts were tested to elucidate the distribution and gradient of V_{OX} across the film. Inert gases such as N_2 and Ar were shown to create V_{OX} near the WO_X surface after high-temperature annealing, while O_2 was shown to annihilate V_{OX}. The V_{OX} concentration and distribution strongly affect the RS characteristics, including switching polarity, current level, and operating voltage. Vacuum tests were in agreement with the understanding that oxygen exchange with ambient and metal WF are both in play.

Calculation of thermal effects in the film during switching was performed; in analog operation, the temperature increase is estimated to be about tens of degrees and in digital operation, hundreds of degrees. Finally, a simple model was postulated where RS is due to the migration of V_{OX} from the TE/WO_X interface upon positive bias on the TE. The knowledge on device/material characteristics gained through these experiments are very useful in optimizing WO_X memristors for various applications each having distinct requirements.

Appendix

XPS study of sputtered WO_X films

Besides RTA-oxidized WO_X films, sputtered WO_X films were also analyzed [14]. The different stoichiometry was achieved by changing the ratio between Ar and O_2 while keeping the total flow rate fixed during sputtering. Figure 3A.1 shows the XPS W4*f* spectra of four different O_2 percentage values.

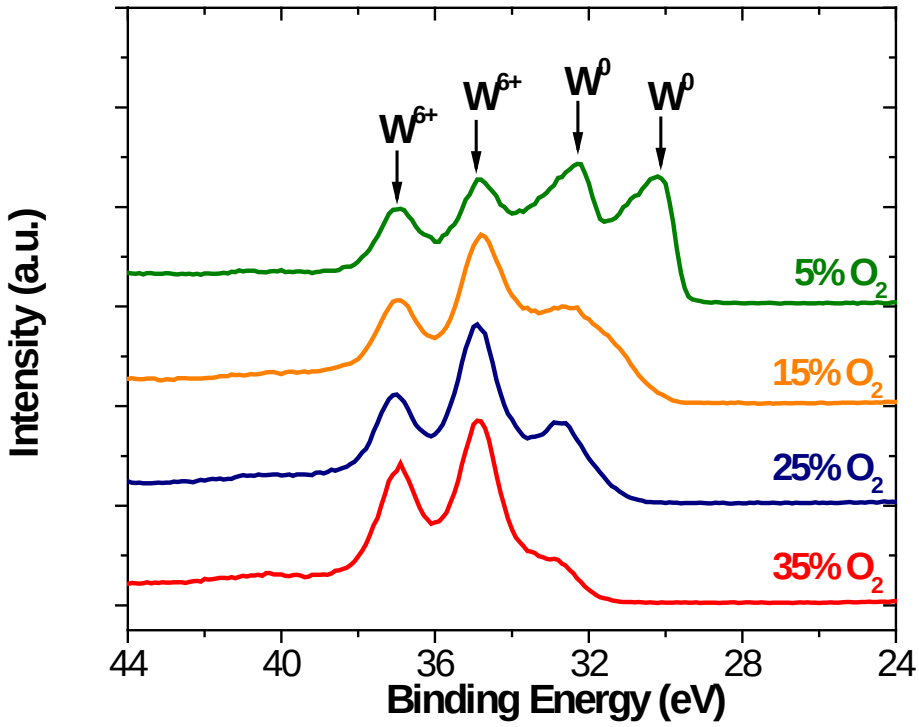


Figure 3A. 1 XPS W4*f* core spectra of sputtered WO_X films prepared under different O_2 flow rate. The 35 eV – 37eV doublet corresponds to the W^{6+} oxidation state and the 30 eV– 32 eV double corresponds to the W^0 oxidation state. There is some offset in the energy spectrum that is not calibrated.

Fabrication of sputtered WO_X devices

The starting substrate was Si/SiO_2 with a NiCr/Pd evaporated film on top as the global BE. A 65 nm PECVD SiO_2 film was deposited at 280 °C for electrical/structural

isolation. Openings to the BE were patterned and etched by EBL and RIE, respectively. The WO_X film was then sputtered using a W target in an oxygen-containing (O_2 : Ar = 1 : 3) chamber at 300 °C. The WO_X thickness was ~ 40 nm. For the annealing study, the WO_X top surface was subjected to different annealing gases, including N_2, Ar, and O_2 at 400 °C for 1 min. Lastly, the TE's were patterned by EBL followed by Pd/Au liftoff. Figure 3A.1 shows the cross-section view of the final structure.

For the metal electrode study, WO_X surface was annealed in O_2 at 400 °C for 1 min. Au, Ni/Au, Al/Au, and Ti/Au TE materials were tested. Au was capped for better resistance to WO_X-etch by RIE.

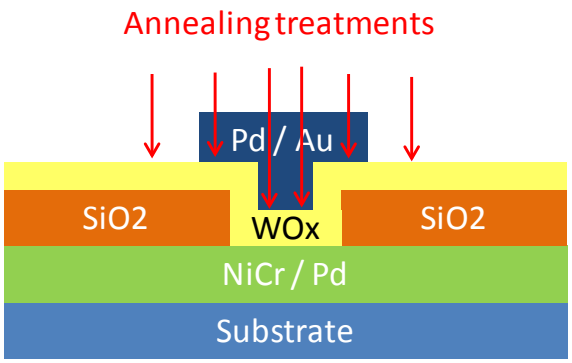


Figure 3A. 2 Final structure of the sputtered Pd/WO_X/Pd devices.

Reference

[1] J. G. Simmons, R. R. Verderber, "New Conduction and Reversible Memory Phenomena in Thin Insulating Films," *Proc. Roy. Soc. A* **301**, 77, 1967.

[2] A. Sawa, "Resistive Switching in Transition Metal Oxides," *Mater. Today* **11**, 28, 2008.

[3] R. Waser, R. Dittmann, G. Staikov, K. Szot, "Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects, and Challenges," *Adv. Mater.* **21**, 2632, 2009.

[4] H. A. Wriedt, "The O-W (Oxygen-Tungsten) System," *J. Phase Equilib.* **10**, 368, 1989.

[5] D. B. Migas, V. L. Shaposhnikov, V. E. Borisenko, "Tungsten Oxides. II. The Metallic Nature of Magnéli Phases," *J. Appl. Phys.* **108**, 093714, 2010.

[6] M. T. Greiner, L. Chai, M. G. Helander, W. M. Tang, Z. H. Lu, "Transition Metal Oxide Work Functions: The Influence of Cation Oxidation State and Oxygen Vacancies," *Adv. Funct. Mater.*, DOI: 10.1002/adfm.201200615.

[7] M. T. Greiner, L. Chai, M. G. Helander, W. M. Tang, Z. H. Lu, "Metal/Metal-Oxide Interfaces: How Metal Contacts Affect the Work Function and Band Structure of MoO_3," *Adv. Funct. Mater.*, DOI: 10.1002/adfm.201200993.

[8] D. S. Shang, L. Shi, J. R. Sun, B. G. Shen, F. Zhuge, R. W. Li, Y. G. Zhao, "Improvement of Reproducible Resistance Switching in Polycrystalline Tungsten Oxide Films by *in situ* Oxygen Annealing," *Appl. Phys. Lett.* **96**, 072103, 2010.

[9] K. P. Biju, X. Liu, M. Siddik, S. Kim, J. Shin, I. Kim, A. Ignatiev, H. Huang, "Resistive Switching Characteristics and Mechanism of Thermally Grown WO_X Thin Films," *J. Appl. Phys.* **110**, 064505, 2011.

[10] R. Sohal, C. Walczyk, P Zaumseil, D. Wolansky, A. Fox, B. Tillack, H. J. Müssig, T. Schroeder, "Thermal Oxidation of Chemical Vapour Deposited Tungsten Layers on Silicon Substrates for Embedded Non-Volatile Memory Application," *Thin Solid Films* **517**, 4534, 2009.

[11] K. McKenna, A. Shluger, "The Interaction of Oxygen Vacancies with Grain Boundaries in Monoclinic HfO_2," Appl. *Phys. Lett.* **95**, 222111, 2009.

[12] S. Santucci, C. Cantalini, M. Crivellari, L. Lozzi, L. Ottaviano, M. Passacantando, "X-Ray Photoemission Spectroscopy and Scanning Tunneling Spectroscopy Study on the Thermal Stability of WO_3 Thin Films," *J. Vac. Sci. Technol. A* **18**, 1077, 2000

[13] K. Masek, J. Libra, T. Skála, M. Cabala, V. Matolin, V. Cháb, K. C. Prince, "SRPES Investigation of Tungsten Oxide in Different Oxidation States," *Surf. Sci.* **600**, 1624, 2006.

[14] E. Griffin, J. Holt, *unpublished*.

[15] H. J. T. Ellingham, "Reducibility of Oxides and Sulphides in Metallurgical Processes," *J. Soc. Chem. Ind.* **63**, 125, 1944.

[16] J. J. Yang, J. P. Strachan, F. Miao, M. X. Zhang, M. D. Pickett, W. Yi, D. A. A. Ohlberg, G. Medeiros-Ribeiro, R. S. Williams, "Metal/TiO_2 Interfaces for Memristive Switches," *Appl. Phys. A – Mater.* **102**, 785, 2011.

[17] N. Zhong, H. Shima, H. Akinaga, "Rectifying Characteristics of Pt/TiO_X/Metal/Pt Controlled by Electronegativity," *Appl. Phys. Lett.* **96**, 042107, 2010.

[18] "CRC Handbook of Chemistry and Physics, 92nd edition," CRC Press, 1977.

[19] http://www.doitpoms.ac.uk/tlplib/ellingham_diagrams/index.php

[20] R. Yang, K. Terabe, T. Tsuruoka, T. Hasegawa, M. Aono, "Oxygen Migration Process in the Interface during Bipolar Resistance Switching Behavior of WO_{3-x}-Based Nanoionics Devices," *Appl. Phys. Lett.* **100**, 231603, 2012.

[21] D. S. Jeong, H. Schroeder, U. Breuer, R. Waser, "Characteristic Electroforming Behavior in Pt/TiO_2/Pt Resistive Switching Cells Depending on Atmosphere," *J. Appl. Phys.* **104**, 123716, 2008.

[22] K. Miyake, H. Kaneko, M. Sano, N. Suedomi, "Physical and Electrochromic Properties of the Amorphous and Crystalline Tungsten Oxide Thick Films Prepared under Reducing Atmosphere," *J. Appl. Phys.* **55**, 2747, 1984.

[23] M. D. Pickett, J. Borghetti, J. J. Yang, G. Medeiros-Ribeiro, R. S. Williams, "Coexistence of Memristance and Negative Differential Resistance in a Nanoscale Metal-Oxide-Metal System," *Adv. Mater.* **23**, 1730, 2011.

[24] M. D. Pickett, R. S. Williams, "Sub-100 fJ and Sub-Nanosecond Thermally Driven Threshold Switching in Niobium Oxide Crosspoint Nanodevices," *Nanotechnol.* **23**, 215202, 2012.

[25] F. Miao, J. P. Strachan, J. J. Yang, M. X. Zhang, I. Goldfarb, A. C. Torrezan, P. Eschbach, R. D. Kelley, G. Medeiros-Ribeiro, R. S. Williams," Anatomy of a Nanoscale Conduction Channel Reveals the Mechanism of a High-Performance Memristor," *Adv. Mater.* **23**, 5633, 2011.

[26] G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, M. Nafria, "Metal Oxide Resistive Memory Switching Mechanism Based on Conductive Filament Properties," *J. Appl. Phys.* **110**, 124518, 2011.

[27] L. L. Y. Chang, B. Phillips, "Phase Relations in Refractory Metal-Oxygen Systems," *J. Am. Ceram. Soc.* **52**, 527, 1969.

[28] A. A. Mohammad, M. Gillet, "Phase Transformations in WO3 Thin Films during Annealing," *Thin Solid Films* **408**, 302, 2002.

[29] G. Dearnaley, A. M. Stoneham, D. B. Morgan, "Electrical Phenomena in Amorphous Oxide Films," *Rep. Prog. Phys.* **33**, 1129, 1970.

[30] H. Wang, Y. Xu, M. Goto, Y. Tanaka, M. Yamazaki, A. Kasahara, M. Tosa, "Thermal Conductivity Measurement of Tungsten Oxide Nanoscale Thin Films," *Mater. Trans.* **47**, 1894, 2006.

Chapter 4
Modeling of Analog Switching in WO_X Memristors

4.1 Introduction

After having much knowledge about WO_X memristors, an immediate need for a device model emerged, not only for better understanding of device dynamics, but also for accommodating the model to different systems. Moreover, this model should facilitate in designing circuits and systems that utilize memristors. There are different levels of approaches for building device models. Material scientists modeled switching using *ab initio* computations to simulate the dynamics of V_{OX} in the metal-oxide crystal structure [1] while chemists modeled switching using electrochemical reaction calculations [2]. As electrical engineers, the approach was to identify the phenomenological equations that capture realistic device physics while maintaining succinctness. Therefore, a compact model was developed which could also be applied to different memristive systems as long as appropriate expressions are chosen.

4.2 Modeling of Analog Switching

Modeling nanoscale memristor switching using Chua's equations was first accomplished by the HP Labs for TiO_X devices [3-5]. The memristor equations are

$$v = M(w,i)i \tag{4.1}$$

$$\frac{dw}{dt} = f(w,i) \tag{4.2}$$

where M is the memristance. The HP model used the conducting region (CR) length as the internal state variable (w) to explain the switching polarity and the I-V nonlinearity; the resistance of the device was determined by the series connection of a doped region and an undoped region separated by a movable boundary, as illustrated in Figure 4.1(a). Equations 4.1 and 4.2 can be expressed as

$$v(t) = (R_{ON}\frac{w(t)}{d} + R_{OFF}(1-\frac{w(t)}{d}))i(t) \tag{4.3}$$

$$\frac{dw(t)}{dt} = \mu_V \frac{R_{ON}}{d} i(t) \tag{4.4}$$

where R_{OFF} (R_{ON}) is the resistance of the undoped (doped) region, w is the thickness of the doped region, d is the total thickness of the film, and μ_V is the average ion mobility. This model is easy to capture but oversimplified. Later, they refined the model by adding nonlinear carrier transport and interface barrier effects to account for field-enhanced mobility and rectification [4-6].

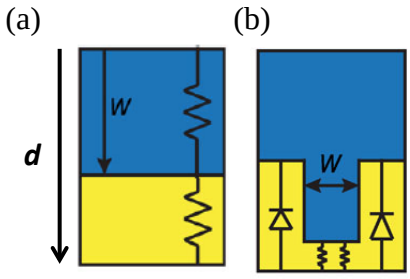


Figure 4.1 Illustrations of using the internal state variable w to represent (a) the CR length and (b) the CR area (or CR width in this 2-D view). Blue and yellow denote the doped and undoped region, respectively.

We note that the HP model describes a "current-controlled" memristor, as the *I-V* and dynamic equations are functions of current. In reality, voltage-control is much more convenient to implement. For a voltage-controlled memristor, the memductance equations are provided here again for convenience.

$$i = G(w, v)v \tag{4.5}$$

$$\frac{dw}{dt} = f(w, v) \tag{4.6}$$

Additionally, the assumption in Equation 4.4 is the linear relation between the ion movement (*dw/dt*) and the ion mobility (μ_V). This is not very realistic as the ion mobility is usually enhanced at high electric fields [6]. An exponential dependence would more closely describe ionic drift

$$\frac{dw}{dt} \propto \sinh(\frac{v / E_0}{d - w}) \tag{4.7}$$

where E_0 is a characteristic field. As *w* increases, (*dw/dt*) follows in a very nonlinear fashion. For simplicity, *d* was set to 1 so that *w* is normalized between 0 and 1.

If the state variable *w* is assumed to be the length of the conductive region, then

$$G(w) = \frac{1}{R_{ON}\frac{w}{d} + R_{OFF}(1 - \frac{w}{d})} \tag{4.8}$$

Plugging in Equations 4.7 and 4.8, the simulated RS behavior was obtained, as shown in Figure 4.2 [7]. Here the step changes resulted from identical voltage sweeps were very nonlinear; the step change was largest when *w* approached 1. These behaviors, although clearly captured the pinched-hysteresis effects, were distinctly different from our experimental observations (Figure 2.3) where every voltage sweep roughly caused a constant conductance change.

(a) (b)

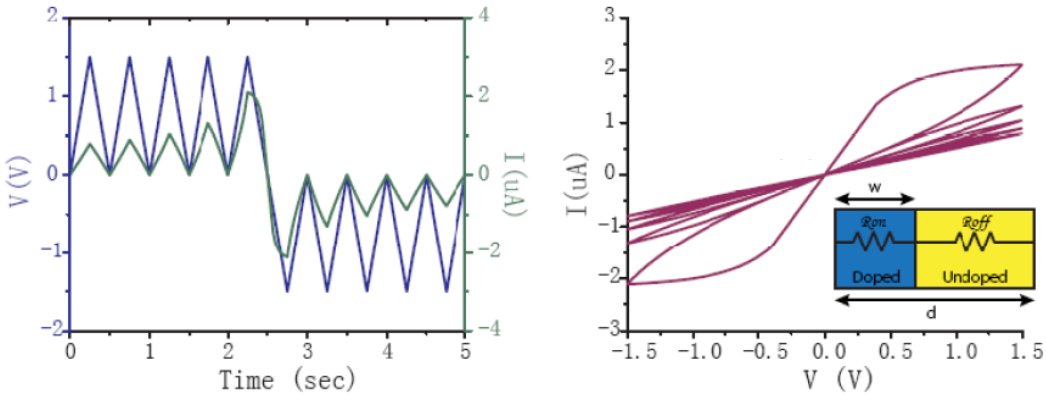


Figure 4.2 Simulation results using *w* as CR length (Equations 4.7 and 4.8). (a) Voltage and current plotted against time. (b) *I-V* curves extracted from (a).

Therefore, we proposed that in our analog memristive devices, *w* can be better expressed as the area index (Figure 4.1(b)). In the Pd/WO_X/W devices, both the WO_X/W and the Pd/WO_X interfaces can form Schottky barriers. As mentioned before, the thermal oxidation process makes the top surface more crystallized than the regions buried below, therefore creating a non-uniform distribution of V_{OX} as-fabricated, with more V_{OX} near the top interface than near the bottom interface. As a result, the top Pd/WO_X interface was assumed to be Ohmic and act as a V_{OX} reservoir, and the switching action occurs near the bottom W/WO_X interface. Upon the application of a positive voltage, V_{OX} drift towards BE. The areas with abundant V_{OX} will result in an Ohmic-like contact dominated by tunneling current. The memristor equations can then be written as

$$i = (1-w)\alpha[1-\exp(-\beta v)] + w\gamma\sinh(\delta v) \tag{4.9}$$

$$\frac{dw}{dt} = \lambda\sinh(\eta v) \tag{4.10}$$

In Equation 4.9, the first term represents the portions at the W/WO_X interface with deficient V_{OX} dominated by a Schottky barrier, and the second term represents the conductive portions at the W/WO_X interface dominated by tunneling due to the accumulation of V_{OX}. The two terms conduct current in parallel with relative weight adjusted by w, so $w = 0$ indicates a fully Schottky conduction and $w = 1$ indicates a fully tunneling conduction. Here α, β, γ, δ, λ, η_1, and η_2 are all positive-valued parameters determined by material properties, for example, Schottky and tunneling barrier heights, depletion width near the Schottky junction, tunneling distance near the Ohmic junction, as well as interface effects. In modeling, they were treated as w-independent fitting parameters. In the dynamic Equation 4.10, (dw/dt) is treated as a function only of the applied bias independent of w. This is reasonable under the assumption that existing CR does not affect the formation of another CR (or the expansion of existing CR) as the CR forms in parallel under voltage-bias. The resulting switching curves are shown in Figure 4.3. Note that this assumption is not exact, so improvement of the model will be covered in Section 4.4.

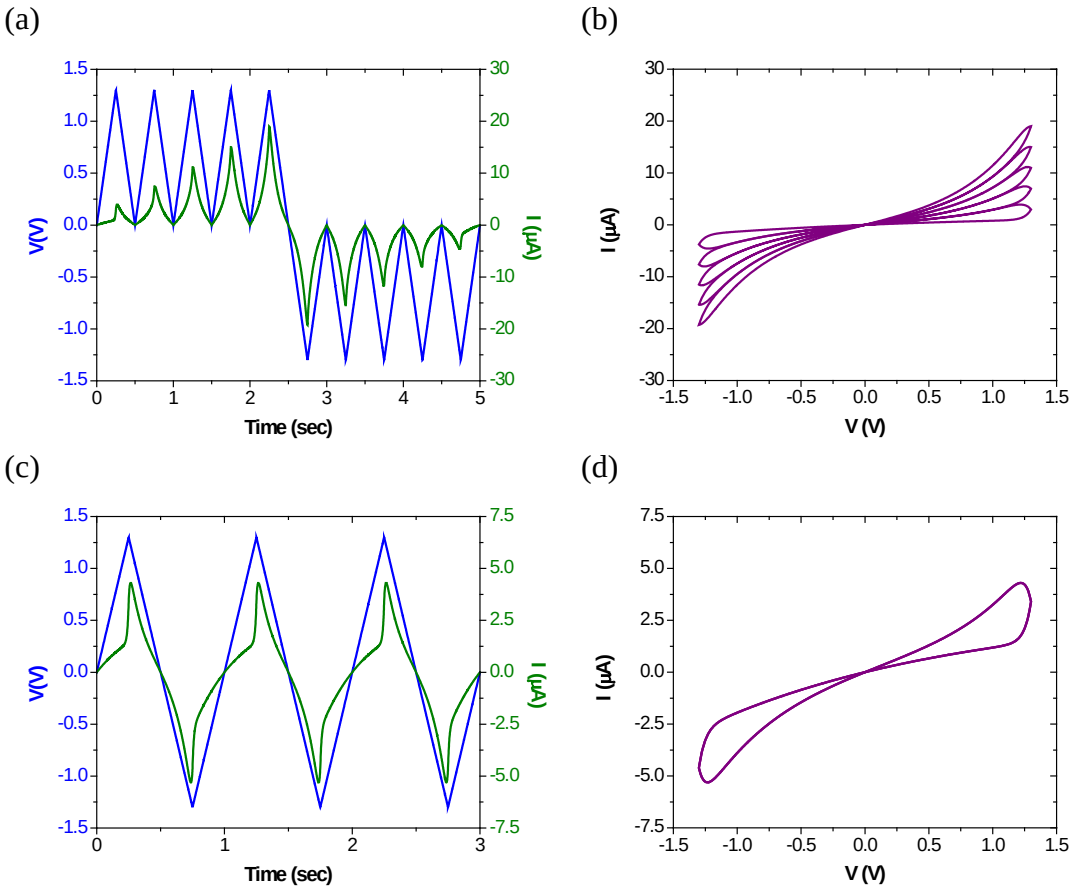


Figure 4.3 Simulation results using w as CR width (Equations 4.9 and 4.10). Voltage and current plotted against time under (a) five consecutive +1.3 V sweeps followed by five consecutive -1.3 V sweeps and (c) three alternating ±1.3 V sweeps. (b), (d) are *I-V* curves extracted from (a), (c), respectively.

4.3 Implementation of Memristor Model in SPICE

As a proof of concept, we used LTspice, a free commercial SPICE (Simulation Program with Integrated Circuit Emphasis) software, for demonstration. In order to implement the rate equation, we needed an integrator. The integrator was realized using a floating capacitor in a sub-circuit which solved Equations 4.9 and 4.10 self-consistently, employing the scheme in [8]. A sample code is provided in Table 4.1.

Table 4.1 LTspice code of the memristor model.

```
*Parameters:
*alpha is prefactor for Schottky barrier
*beta is exponent for Schottky barrier
*gamma is prefactor for tunneling
*delta is exponent for tunnling
**********************************************************************************************
.subckt analog 1 2 params:
+ alpha=2e-6 beta=0.5 gamma=4e-6 delta=2 wmax=1 wmin=0
*State variable:
.param lambda=1 eta1=1e-9 eta2=18
.param cp={1}
Cpvar w 0 {cp}
Gx 0 w value={trunc(V(1,2),cp*V(w))*lambda*(eta1*sinh(eta2*V(1,2)))}
.ic V(w) = 0.0
**********************************************************************************************
*Window function definition
.func sign2(var) {(sgn(var)+1)/2}
.func trunc(var1,var2) {sign2(var1)*sign2(wmax-var2)+sign2(-var1)*sign2(var2-wmin)}
**********************************************************************************************
*Output:
Gw 1 2 value={(1-cp*V(w))*alpha*(1-exp(-
beta*V(1,2)))+(cp*V(w))*gamma*sinh(delta*V(1,2))}
.ENDS memristor
```

4.4 Optimization of Memristor Model

The first discrepancy noticed in the ideal model was retention, or, the overlap of *I-V* curves in Figure 2.3(a). To account for the natural V_{OX} diffusion, a decay term was added to Equation 4.10,

$$\frac{dw}{dt} = \lambda \sinh(\eta v) - \frac{w}{\tau} \tag{4.11}$$

which had a time constant τ on the order of a few seconds. This decay term corresponds to the natural diffusion of V_{OX} from the conductive regions after programming. The resulting curves are shown in Figure 4.4. It can be clearly seen in Figure 4.4(b) that on the positive side the *I-V* curves overlapped with preceding/subsequent sweeps and on the negative side the minimum conductance was reached quickly after the third sweep.

Effects from the fourth and fifth negative sweeps were negligible because the hard boundary for w was reached (i.e. w = 0). This issue can be solved by imposing an appropriate window function.

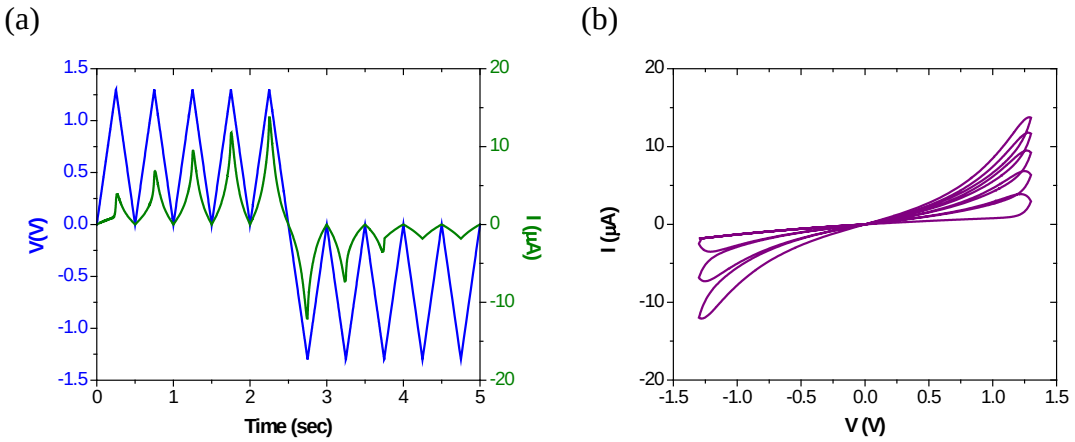


Figure 4.4 Simulation results including V_{OX} diffusion effects, in which Equation 4.10 was replaced with Equation 4.11.

Window functions are commonly used in simulation for smoothing purposes as the state variables approaches the boundaries. In [8-9], a window function (Equation 4.12) was used to reduce the rate of change of w whenever it was approaching its boundaries (i.e. w = 0 and 1), written as

$$h(x) = 1 - (2x - 1)^{2p} \tag{4.12}$$

where p is a positive integer. However, a slightly different behavior was observed in our experiments such that Equation 4.12 seemed inaccurate. In the WO_X devices, the first few operations subsequent to a polarity change were always most effective. In other words, in a series of positive bias sweeps or pulses, the first few cycles were effectively stronger than later ones; the situation was similar with negative operations. Thus the rate equation was modified to be proportional to (1-w) under positive bias and to w under

negative bias. Alternatively, one can think of this window function as the slowing down of (*dw/dt*) as *w* is reaching its final value of that particular operation. Then the switching curves would look like Figure 4.5. From Figure 4.5(b), the decremented effectiveness of identical sweeps is seen. As will be discussed in Chapter 6, this behavior shares a common signature with the famous BCM (Bienenstock-Cooper-Munro) theory for biological systems [10].

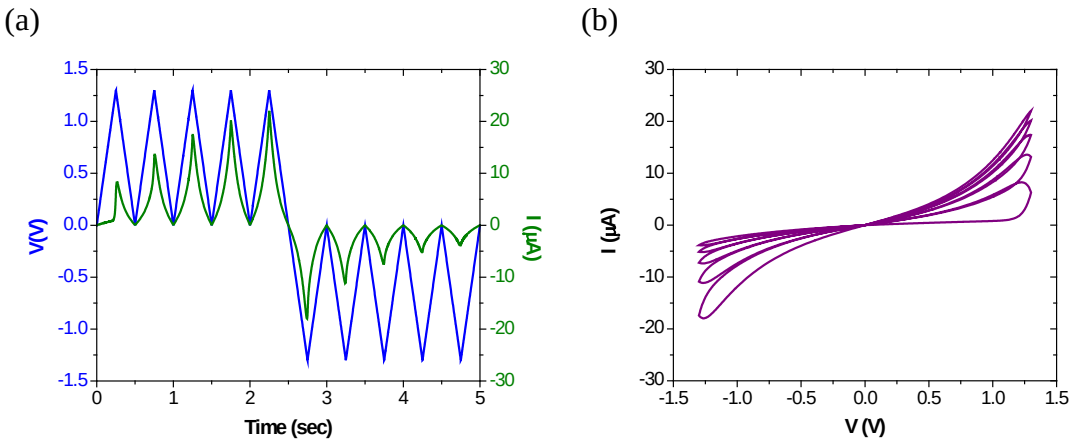


Figure 4.5 Simulation results adding a window function. (*dw/dt*) was modified to be proportional to (1-*w*) under positive bias and to *w* under negative bias.

Not only did the model support DC operations with this model, pulse operations could also be reproduced accurately. Figure 4.6 shows the progression of improvements being made to the model to capture pulse operation more closely. By comparing these figures with Figure 2.4, we see that much optimization has been made to allow the compact model agree well with experiments.

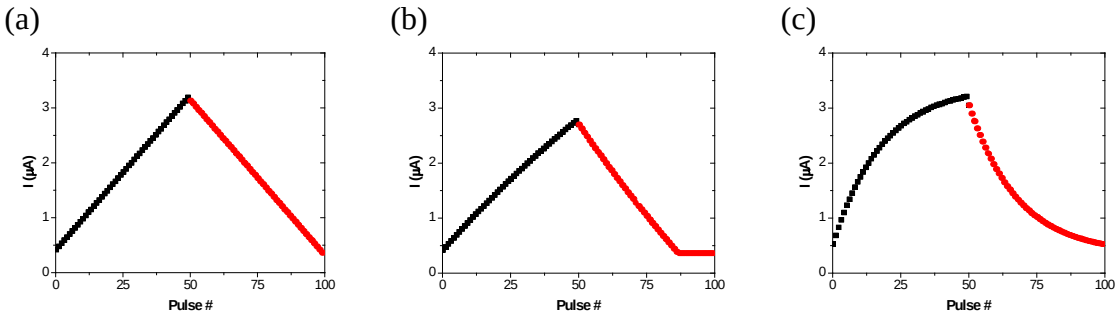


Figure 4.6 Simulation results of pulse operation. (a) Simple modeling using *w* as CR width (area) → (b) added a diffusion term → (c) added a window function. P: +1.4 V, 400 μs, D: -1.4V, 400 μs, and read: 0.4V, 3 ms.

4.5 Verification of Memristor Model

The optimized model was then evaluated with actual experimental data, and indeed, the model captured the switching characteristics, including voltage, current, and time scales. Figure 4.7 and 4.8 show the fitting results for DC and pulse operations, respectively. These results indicate that this compact, phenomenological memristor model is convenient and realistic.

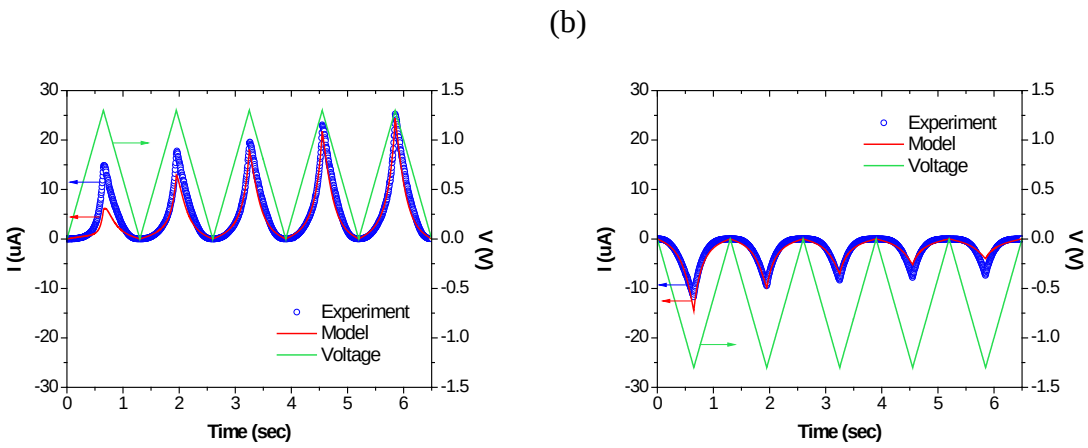


Figure 4.7 Experimental DC data fitted by the memristor model. (a) Five positive sweeps followed by (b) five negative sweeps.

(a) (b)

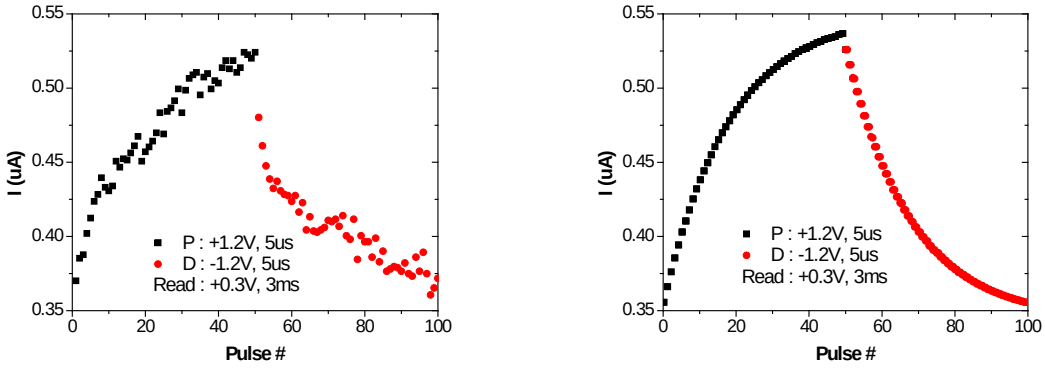


Figure 4.8 Experimental pulse data fitted by the memristor model. (a) Experimental and (b) fitting of identical pulsing conditions.

4.6 Conclusion

Device models, on the one hand, help researchers understand switching dynamics in their devices, and on the other hand, facilitate circuit design for various applications. Here a phenomenological model based on the memristor equations was developed, with the incorporation of an internal state variable that weighs the conduction between Schottky emission and tunneling. The model is realistic, compact, and flexible. Most importantly, RS mechanisms were physically implemented in circuit simulators with further optimization, making the modeling of different memristive devices and simulation of circuit-level designs much more convenient.

Reference

[1] D. B. Migas, V. L. Shaposhnikov, V. E. Borisenko, "Tungsten Oxides. II. The Metallic Nature of Magnéli Phases," *J. Appl. Phys.* **108**, 093714, 2010.

[2] D. S. Jeong, H. Schroeder, R. Waser, "Mechanism for Bipolar Switching in a Pt/TiO_2/Pt Resistive Switching Cell," *Phys. Rev. B* **79**, 195317, 2009.

[3] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, "The Missing Memristor Found," *Nature* **453**, 80, 2008.

[4] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, R. S. Williams, "Memristive Switching Mechanism for Metal/Oxide/Metal Nanodevices," *Nat. Nanotechnol.* **3**, 429, 2008.

[5] D. B. Strukov, J. L. Borghetti, R. S. Williams, "Coupled Ionic and Electronic Transport Model for Thin-Film Semiconductor Memristive Behavior," *Small* **5**, 1058, 2009.

[6] D. B. Strukov, R. S. Williams, "Exponential Ionic Drift: Fast Switching and Low Volatility of Thin-Film Memristors," *Appl. Phys. A – Mater.* **94**, 515, 2009.

[7] T. Chang, S. H. Jo, K. H. Kim, P. Sheridan, S. Gaba, W. Lu, "Synaptic Behaviors and Modeling of A Metal Oxide Memristive Device," *Appl. Phys. A – Mater.* **102**, 857, 2011.

[8] Z. Biolek, D. Biolek, V. Biolkova, "SPICE Model of Memristor with Nonlinear Dopant Drift," *Radioengineering* **18**, 210, 2009.

[9] Y. N. Joglekar, S. J. Wolf, "The Elusive Memristor: Properties of Basic Electrical Circuits," *Eur. J. Phys.* **30**, 661, 2009.

[10] E. L. Bienenstock, L. N. Cooper, P. W. Munro, "Theory for the Development of Neuron Selectivity: Orientation Specificity and Binocular Interaction in Visual Cortex," *J. Neurosci.* **2**, 32, 1982.

Chapter 5
Memory Formation and Memory Loss

5.1 Introduction

In biological systems, “memory” is an essential building block for learning. Unlike modern semiconductor memory systems, needless to say, human memory is by no means eternal. Yet, forgetfulness is not always a disadvantage since it releases memory storage for more important or more frequently accessed pieces of information and is thought to be necessary for individuals to adapt to new environments. Eventually, only memories that are of significance are transformed from short-term memory (STM) into long-term memory (LTM) through repeated stimulation.

Efforts to understand STM and LTM are substantial. On the one hand, having studied memory retention for over a century, psychologists have identified several possible causes for forgetting, including memory decay, interference, and aging [1-3]. On the other hand, through anatomy and electrical recordings, neuroscientists have discovered many important mechanisms, although many physiological processes are still under debate [4-5]. No matter what the exact mechanisms are, forgetting is an indisputable fact. Simply speaking, short memory retention is attributed to transient activation or transmission of certain chemicals, which could be ions, transmitters, receptors, etc.

It was previously mentioned that the retention of WO_X memristors is short due to the high mobility of V_{OX}. Short retention is generally not preferred for traditional memory applications, but it actually mimics the decay of ion flows at synapses. In this chapter, retention study is carried out and how "memory decay" in WO_X memristors resembles forgetting in biological systems is explored.

5.2 Stretched-Exponential Decay

To characterize the decay behavior of WO_X memristors the retention curves were recorded using the standard scheme for retention test in modern memory technology: first set the device to the desired state, and then apply a periodic train of low voltage pulses to read out the state of the device without disturbance. Figure 5.1(a) shows the retention curve of a WO_X device, where continuous decay of the state is obvious with a characteristic decay time of ~ minutes.

The retention loss in the WO_X memristors can be explained by the spontaneous diffusion of V_{OX} in the WO_X film [6]. Quantitatively, it is suitable to model retention loss with a stretched-exponential function (SEF) [7-9] which has been widely used to describe electronic or structural relaxation in disordered materials such as glass, polymers, or other dielectrics. For example, hydrogen diffusion in α-Si [10] and defect diffusion and hopping transport in complex condensed-matter systems [11] have been shown to agree well with this model. SEF, also known as the Kohlrausch law, is written as

$$\phi(t) = I_0 \exp[-(\frac{t}{\tau})^{\beta}] \tag{5.1}$$

where $\varphi(t)$ is the relaxation function, τ is the characteristic relaxation time, I_0 is a prefactor, and β is the stretch index ranging between 0 and 1. The stretched exponential behavior originates from the wide distribution of activation energies and the associated wide spread of relaxation times among different relaxation processes in disordered systems. Thus, τ and β in Equation 5.1 jointly account for the collective behavior of all possible relaxation processes for V_{OX} in the memristor system. The data (dots) were reasonably fitted by the SEF (curve), as shown in Figure 5.1(a).

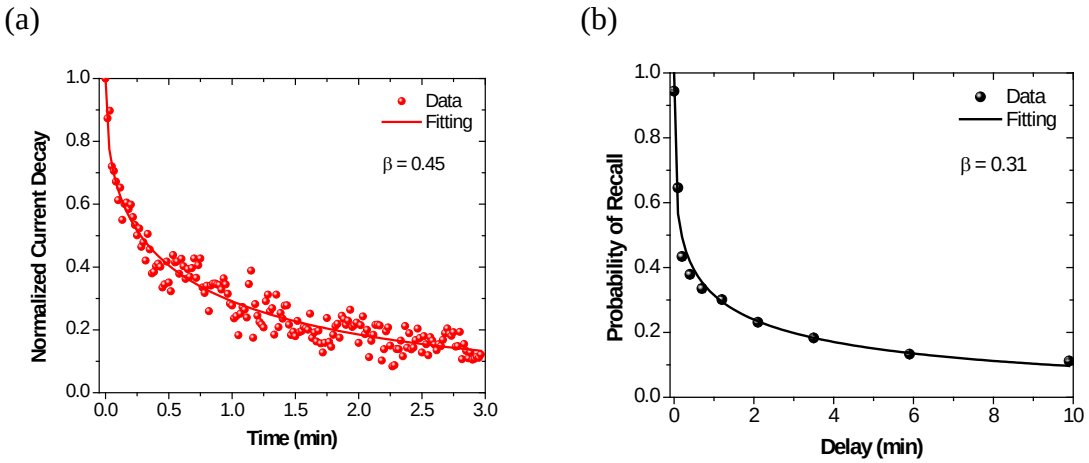


Figure 5.1 Memristor retention curve vs. human forgetting curve. (a) Retention curve of a WO_X memristor. The initial state was set by twenty-five +1.3 V, 400 μs pulses applied at 60 ms constant interval and the resistive states were read by +0.5 V, 3 ms pulses at 1 sec constant interval. (b) Forgetting curve of human memory replicated and refitted from [12].

Examining Equation 5.1, one expects a steep drop followed by a gradual decline. This behavior coincides with human memory loss obtained in psychological studies, described as "a rapid initial decline is usually followed by a long, slow decay" [13]. Figure 5.1(b) shows the experimental data (dots) describing human memory loss, which were also fitted by the SEF (curve). Here, probability of recall is a measure of how successful a person can recall something remembered in the past after a period of time.

From the psychological standpoint, human memory trace can be viewed as a chain consisting of numerous links, where one broken link leads to the failure of the entire chain and causes the memory to become inaccessible [1]. It is then natural to describe memory loss with the Weibull distribution, a statistics function widely used in failure analysis (or equivalently, survival analysis) [14-15]. Mathematically, the cumulative Weibull distribution is just a complementary function of the SEF (see Appendix). In fact, there have been examples of psychologists characterizing memory retention with the exponential power function, a synonym for SEF [1, 16].

The apparent agreement with psychological data is so far purely phenomenological. Physically, from the neurobiological point of view, memory is thought to be closely associated with synaptic weights, i.e. the strength of synaptic connections. Strengthening and weakening (i.e. the plasticity) of synaptic weights are found to be governed by the concentrations of ionic species (e.g., Ca^{2+}, Na^{+}, Mg^{2+}, and K^{+}) which activate/inhibit the release of neurotransmitters and activation of receptors with certain timing constraints [17]. Hence, one could then argue that memory decay involves a multi-time constant relaxation process resulting in short-term plasticity (STP) in neurobiological terms.

It needs to be noted that the actual mechanisms for memory loss are not clear yet, and describing the various complex biological processes using a single equation is always grossly oversimplified. However, if one assumes the picture that memory loss can be explained macroscopically by the failure of memory trace networks and microscopically by the synaptic modifications with different time scales, it is then perhaps not surprising

that both memory loss and memristor retention loss can be described to first order by the same mathematical equation (i.e., Equation 5.1).

In the next section, the phenomenon in which the memristor retention is enhanced by repeated stimulations and how this retention-enhancing process is analogous to the transition from short-term memory to long-term memory in biological systems will be discussed.

5.3 Memory Enhancement Effect

5.3.1 Processes in Biological Systems

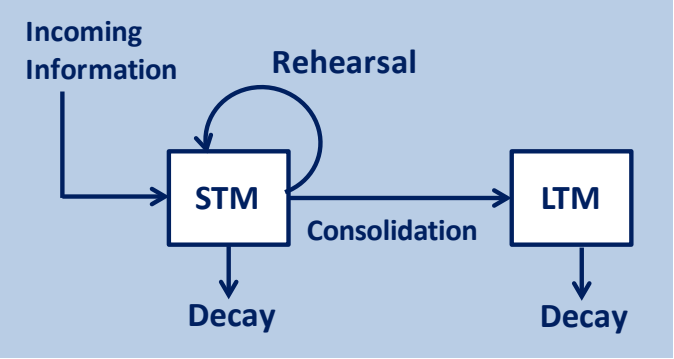


Figure 5.2 Multistore memory model by Shiffrin and Atkinson [16].

In biological systems, STM generally lasts from seconds to tens of minutes; in contrast, LTM lasts from a few hours to days or weeks, sometimes even to a lifetime [18]. Figure 5.2 shows a simplified illustration of the multistore memory model [19]: STM can only be sustained by constantly rehearsing the same stimulus, while LTM, despite the presence of natural forgetting, can be maintained for a much longer period of time without subsequent stimuli. The transition from STM to LTM is also through repetitions

(as in rehearsal) but in this case, is a much more intricate process involving many molecular mechanisms and structural changes at various cellular sites/levels [20]. This entire, complicated process is termed consolidation [21].

More strictly, memory consolidation in biological systems is a complex process comprising synaptic consolidation, system consolidation, and reconsolidation [22]. The most important physiological mechanism underlying synaptic consolidation is long-term potentiation (LTP) through which strong and long-lasting synaptic connections are formed [23]. Whereas synaptic consolidation takes place from minutes to hours after a learning or stimulation event, system consolidation occurs days to weeks after the event, and during this period of time brain circuits are reorganized so that memories become stable within the system. However, memories become labile and unstable after recalls, and reconsolidation is necessary to stabilize, modify, and strengthen LTM. Oftentimes, the formation of STM (rehearsal) and LTM (consolidation) involves not only molecular and cellular reactions, but also cognitive and psychological behaviors [2-3].

In the next session, we attempt to demonstrate Based on an analogy between the memory enhancement effect observed in memristors and the STM-to-LTM transition in biological system. It should be noted that this analogy is mostly phenomenological and by no means forming strict one-to-one correlation with biological systems at the molecular or cellular level.

5.3.2 Demonstrations in WO_X Memristors

According to the multistore model, the key to building up memory is *repetition*. In this experiment, we demonstrate the effects of repetition on memristors. Stimuli were

realized as identical voltage pulses with fixed height, width, and pulse-to-pulse interval. Different numbers of stimuli (N = 5 to N = 40 in steps of 5) were applied to the same memristor starting from the same initial state, and retention curves were recorded right after the last stimulus in each stimulation series. Figure 5.3 shows results obtained from tests after 5, 10, ..., 35, and 40 stimuli, along with fittings using Equation 5.1. Both the retention time (represented as τ in Equation 5.1) and the synaptic weight (represented as I_0 in Equation 5.1) can then be obtained through SEF fitting.

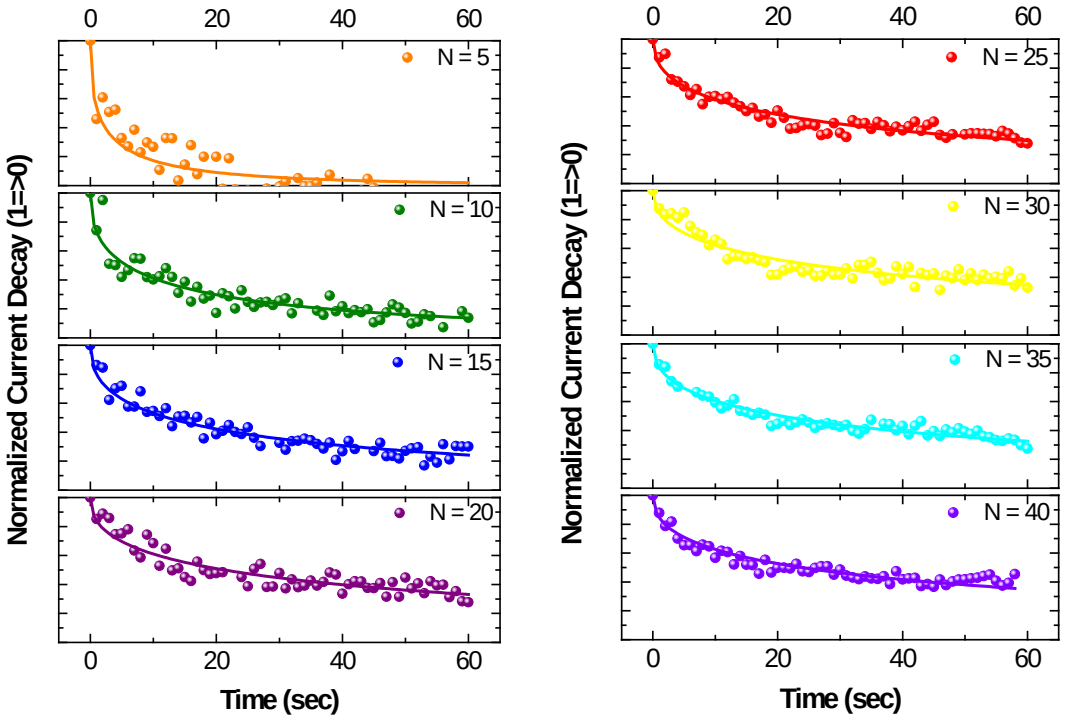


Figure 5.3 Memory retention data (dots) recorded after different numbers of identical stimuli and fitted curves using the SEF (solid lines). Each set of data is normalized by a prefactor I_0 to fall between 1 and 0.

Figure 5.4(a) plots the relationship between the retention time and number of stimuli. A noticeable improvement in τ with repetitive stimulation can be observed.

Overall, τ is increased by approximately 20-fold and slightly saturates beyond N = 20, suggesting two distinct memory regimes in the WO_X memristor, in which one regime with short τ (≈ few seconds) and sensitive to additional stimulations, and another regime with much longer τ (≈ minutes) and relatively insensitive to additional stimulations. Also plotted in Figure 5.4(a) is the synaptic weight (I_0). A similar trend can be found for τ and I_0 with increasing N.

(a) (b)

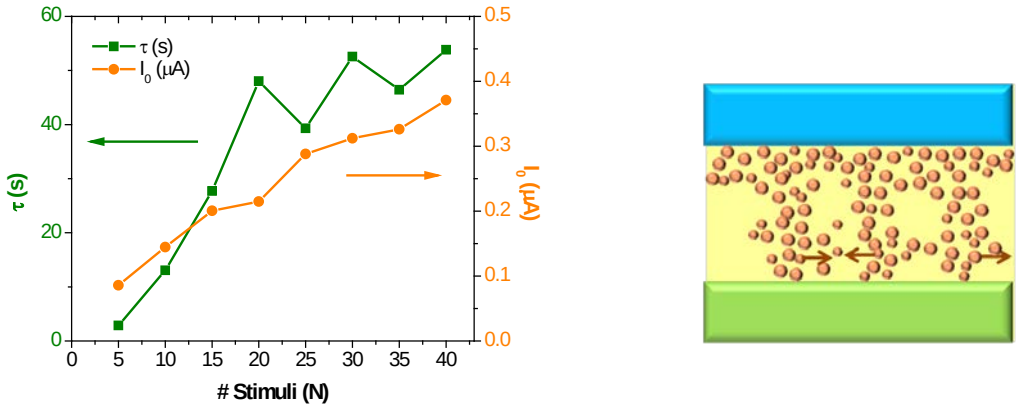


Figure 5.4 STM-to-LTM transition in a WO_X memristor. (a) Characteristic relaxation time (τ) and the prefactor (I_0) obtained through the fitting in Figure 5.3, plotted with respect to the number of stimuli (N). (b) Schematic of the structural change to the memristor corresponding to the improved retention. After repeated stimulation, there are sufficient V_{OX} in the switching layer that lateral diffusion effectively cancels out, resulting in much improved retention along with the increase in conductance (synaptic weight).

The two memory regimes and the transition between them in the memristor can be schematically explained with the aid of Figure 5.4(b). With the addition of repeated stimulation, a higher concentration of V_{OX} is moved into the switching layer, and the lateral diffusion of V_{OX} eventually balances each other out leading to a much lower probability of breakage of the conducting paths. Alternatively, if ρ is the critical V_{OX} density to maintain a conducting path, it is much easier to reach ρ (or much harder to

drop below ρ) if the overall average concentration of V_{OX} is high. Hence, the retention time improves along with the memristor conductance until a significant amount of V_{OX} is stabilized in the film, where the retention time and the conductance saturate.

Interestingly, the process in Figure 5.4(b) is comparable to the scenario in biological systems. Formation, stabilization, and persistence of LTM are supported by experimental evidence of the growth of new synaptic connections and dendrite size/shape change, adding more pathways for synaptic transmission [20]. LTM still fades with time, indicating that synaptic connections retract/shrink with time, but at a much slower pace than the decay during STM. Consequently, these key attributes of the STM-to-LTM transition process, including the significant increase in memory retention after repeated stimulation and the resulting structural change, have been demonstrated in the WO_X memristors.

5.4 Conclusion

The memristor retention curves were obtained experimentally for different conductance states as a result of repetitive stimulation. These curves showed a typical decaying characteristic that could be fitted by the SEF. Macroscopically, the SEF can be regarded as the survival function of memory traces, and microscopically, the SEF represents a relaxation process with multiple time constants. Memory enhancement effect analogous to the STM-to-LTM transition in biological systems was demonstrated in individual WO_X memristors. Upon repeated stimulations, not only was the memristor conductance was increased, but the memristor retention time was also enhanced. Attributed to the higher concentration of V_{OX} migrated into the switching layer during

repetitive stimulation, the lateral diffusion of V_{OX} eventually balances each other out, resulting in a much lower probability of breaking the conducting paths. Similar phenomena were reported in [24], although different material system, fitting functions, and testing protocols were employed.

Appendix

The Weibull distribution

The two-parameter probability density function (*pdf*) of a Weibull distribution is [25]

$$f(x,\lambda,k) = \begin{cases} \frac{k}{x}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases} \quad \text{(5A. 1)}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The corresponding cumulative density function (*cdf*) is

$$F(x;k,\lambda) = 1 - e^{-(x/\lambda)^k} \quad \text{(5A. 2)}$$

for $x \geq 0$, or else $F(x; k, \lambda) = 0$.

From Equations 5A.1 and 5A.2, the survival function (i.e. the complementary *cdf* of Weibull) becomes

$$S(x) = e^{-(x/\lambda)^k} \quad \text{(5A. 3)}$$

which has the form of a stretched-exponential function.

Reference

[1] D. C. Rubin, A. E. Wenzel, "One Hundred Years of Forgetting: A Quantitative Description of Retention," *Psychol. Rev.* **103**, 734, 1996.

[2] S. D. Sala (Ed); *Forgetting*, Psychology Press, 2010.
[3] J. D. Sweatt; *Mechanisms of Memory*, 2nd ed; Academic, 2010.
[4] J. W. Rudy, *The Neurobiology of Learning and Memory*; Sinauer Associates, Inc. Publishers, 2008.
[5] W. M. Cowan, Ed. *Synapses*; Johns Hopkins University Press, 2001.
[6] T. Chang, S. H. Jo, W. Lu, "Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor," *ACS Nano* **5**, 7669, 2011.
[7] H. Scher, M. F. Shlesinger, J. T. Bendler, "Time-Scale Invariance in Transport and Relaxation," *Phys. Today* **44**, 26, 1991.
[8] J. C. Phillips, "Stretched Exponential Relaxation in Molecular and Electronic Glasses," *Rep. Prog. Phys.* **59**, 1133, 1996.
[9] J. Klafter, M. F. Shlesinger, "On the Relationship among Three Theories of Relaxation in Disordered Systems," *Proc. Natl. Acad. Sci. USA* **83**, 848, 1986.
[10] J. Kakalios, R. A. Street, W. B. Jackson, "Stretched-Exponential Relaxation Arising from Dispersive Diffusion of Hydrogen in Amorphous Silicon," *Phys. Rev. Lett.* **59**, 1037, 1987.
[11] B. Sturman, E. Podivilov, "Origin of Stretched Exponential Relaxation for Hopping-Transport Models," *Phys. Rev. Lett.* **91**, 176602, 2003.
[12] D. C. Rubin, S. Hinton, A. Wenzel, "The Precise Time Course of Retention," *J. Exp. Psychol. Learn.* **25**, 1161, 1999.
[13] J. T. Wixted, E. B. Ebbesen, "On the Form of Forgetting," *Psychol. Sci.* **2**, 409, 1991.
[14] S. B. Yuste, G. Oshanin, K. Lindenberg, O. Bénichou, J. Klafter, "Survival Probability of a Particle in a Sea of Mobile Traps: A Tale of Tails," *Phys. Rev. E* **78**, 021105, 2008.
[15] J. Nelson, "Continuous-Time Random-Walk Model of Electron Transport in Nanocrystalline TiO_2 Electrodes," *Phys. Rev. B* **59**, 374, 1999.
[16] W. A. Wickelgren, "Trace Resistance and the Decay of Long-Term Memory," *J. Math. Psychol.* **9**, 418, 1972.
[17] R. S. Zucker, W. G. Regehr, "Short-Term Synaptic Plasticity," *Annu. Rev. Physiol.* **64**, 355, 2002.
[18] E. R. Kandel, "The Molecular Biology of Memory Storage: A Dialogue between Genes and Synapses," *Science* **294**, 1030, 2001.
[19] R. M. Shiffrin, R. C. Atkinson, "Storage and Retrieval Processes in Long-Term Memory," *Psychol. Rev.* **76**, 179, 1969.
[20] R. Lamprecht, J. LeDoux, "Structural Plasticity and Memory," *Nat. Rev. Neurosci.* **5**, 45, 2004.
[21] J. L. McGaugh, "Memory – A Century of Consolidation," *Science* **287**, 248, 2000.

[22] Y. Dudai, “The Neurobiology of Consolidations, or, How Stable is the Engram?” *Annu. Rev. Psychol.* **55**, 51, 2004.

[23] T. V. P. Bliss, G. L. Collingridge, “A Synaptic Model of Memory: Long-Term Potentiation in the Hippocampus,” *Nature* **361**, 31, 1993.

[24] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, “Short-Term Plasticity and Long-Term Potentiation Mimicked in Single Inorganic Synapses,” *Nat. Mater.* **10**, 591, 2011.

[25] D. K. Schroder, Semiconductor Material and Device Characterization, 3rd ed.; IEEE Press, 2006.

Chapter 6
Synaptic Plasticity and Learning Rules

6.1 Introduction

In Chapter 5, the critical role of repetitive stimulation in establishing memory retention in WO_X memristors was emphasized. Memory-keeping is only a single function of synapses. The exceptional characteristic of synapses is their learning ability, that is, their "plasticity" [1]. Abiding by certain learning rules, the synaptic weight can be bi-directionally modulated in an adaptive manner. In other words, learning is a process which decides on the subsequent functions of the neuromorphic system.

Biologically, the weight of a synapse is jointly determined by the firing patterns of both the pre-synaptic and post-synaptic neurons connecting to it. The synapse could either be potentiated or depressed depending on the neuron spike profiles, transmission of ions, activation of receptors, and many other factors. Neurobiologists have discovered several fundamental learning rules, including rate-dependent synaptic plasticity, timing-dependent synaptic plasticity, and cooperativity [2]. While these learning rules may vary according to the specific type and location of a synapse, this study attempts to focus on the generic features of each rule. In this chapter, the characteristics of each learning rule and the memristor-based implementations of these rules will be presented.

6.2 Rate-Dependent Synaptic Plasticity

As its name suggests, the rate (i.e. frequency) of the action potentials determines the direction and the magnitude of the synaptic weight change. One of the early experimental observations was reported by Magleby and colleagues [3]. Post-tenanic potentiation (PTP) was studied by stimulating the frog neuromuscular junction (NMJ) at different rates, from which they identified at least three distinct components; each component had its own rate-dependency and decay time constant, so the overall process could be phenomenologically considered as a dispersed system described by the SEF. Due to these decaying components, the magnitude of facilitation after repetitive stimulation exhibited a strong dependence on the stimulation rate. In another study, paired-pulse facilitation (PPF) was studied in the synapses between the granule cells and Purkinje cells in rat cerebellar slices [4]; the amount of facilitation was inversely correlated with the paired-pulse interval and an elevation in responsiveness for the second pulse was observed when the two pulses were less than 1 sec apart. Below we show that the WO_X memristors exhibit a similar rate-dependency at comparable time scales.

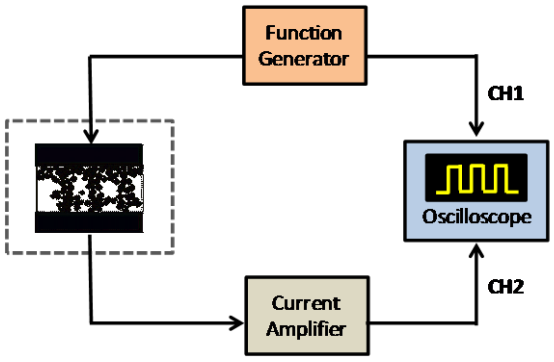


Figure 6.1 Setup for measurements in Figure 6.2. The function generator used was Tektronix AFG 3101, and the oscilloscope used was Tektronix TDS 3032B.

The initial measurements were obtained using an arbitrary function generator and a dual-channel oscilloscope; the setup is illustrated in Figure 6.1. The voltage pulses used here were short in duration so the oscilloscope was not resolving the data points near the peaks very well. To circumvent this, a low DC bias was imposed to read out the state of the device. Figure 6.2(a) shows that at low bias the device state did not drift for at least 2 sec, indicating minimal read disturb – this fact is consistent with the observation that memristor switching is a strong non-linear function of voltage (Equation 4.10). Thus a 0.3 V DC bias for reading out the state and programming pulse trains at 300 ms, 200 ms, and 100 ms intervals were applied. The results are shown in the top panels of Figure 6.2(b), (c), and (d), respectively. For the WO_X memristors, upon the application of trains of positive voltage pulses at different intervals, competing effects between stimulation and decay is anticipated. Indeed, the data show that upon the application of each stimulation pulse, the device conductance was first enhanced, followed by a decay due to spontaneous V_{OX} diffusion. However, when the time interval between the stimulation was sufficiently short, an overall increase in the memristor conductance was observed despite the spontaneous decay, because the idle time between the stimulation pulses was not long enough for the memristor to relax to its initial state, resulting in a net conductance increase. Additionally, a strong trend of the effectiveness of the programming pulse train versus the interval time between pulses can be found; shorter intervals lead to much more effective programming for pulse trains with identical number of pulses, pulse width, and pulse height.

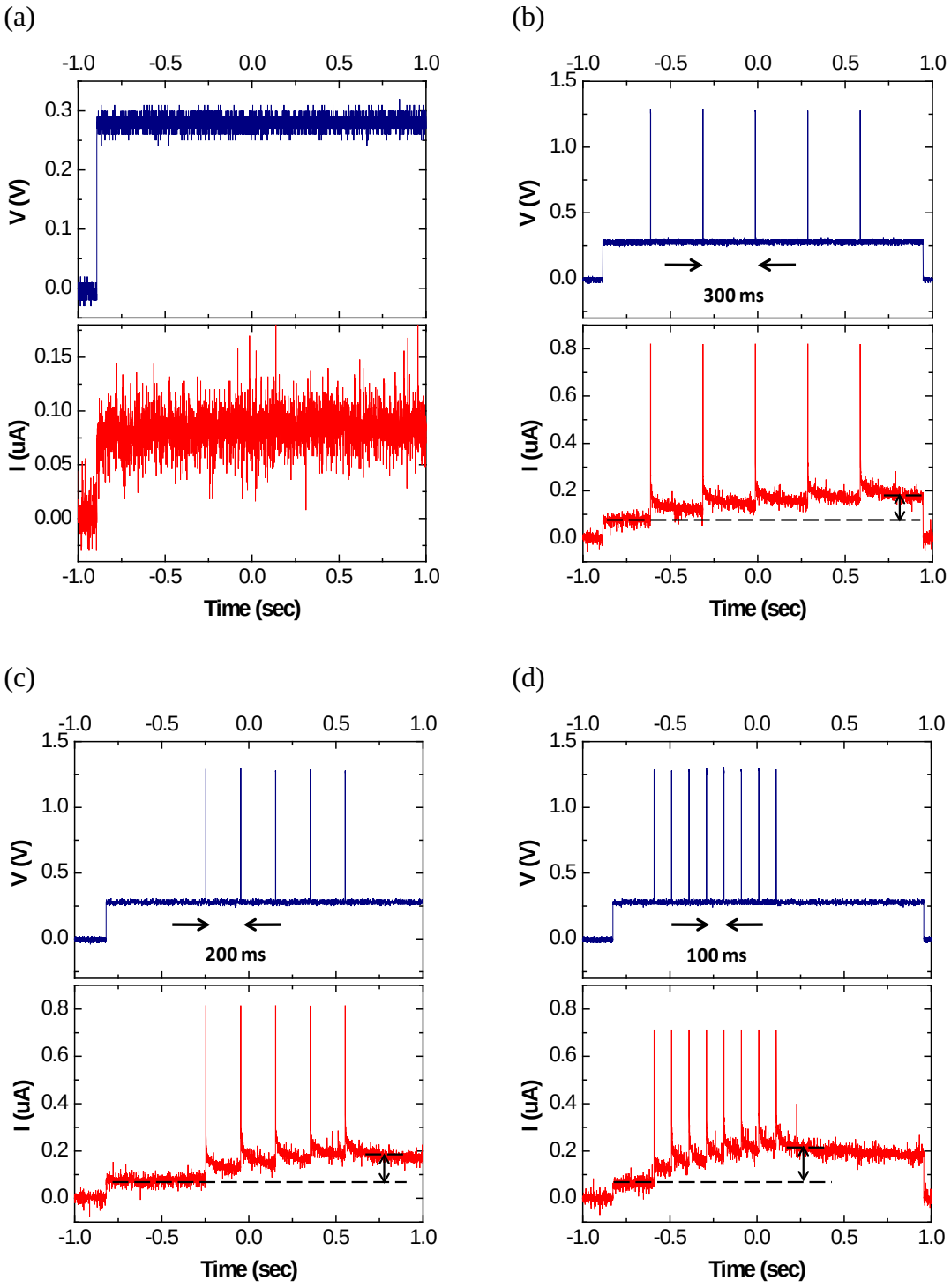


Figure 6.2 Competing effects between stimulation and decay. (a) A 0.3 V DC bias did not disturb the device state. +1.3 V, 1 ms pulses were applied at (b) 300 ms, (c) 200 ms, and (d) 100 ms intervals [5].

To systematically examine the effect of stimulation rate on the memristor, custom LabView and LabWindows programs were used to generate the pulses and record the data replacing function generators and oscilloscopes for better flexibility and easier data collection. Again identical pulses serving as stimuli were applied to the memristor, with the number of stimuli fixed at N = 10 and the interval between stimuli (Δt) varied from 15 ms to 10 sec. This configuration ensured that the total flux (i.e., time integral of the applied voltage) applied to the memristor remains fixed, with the only varying parameter being the stimulation rate. A sample stimulation waveform is illustrated in Figure 6.3.

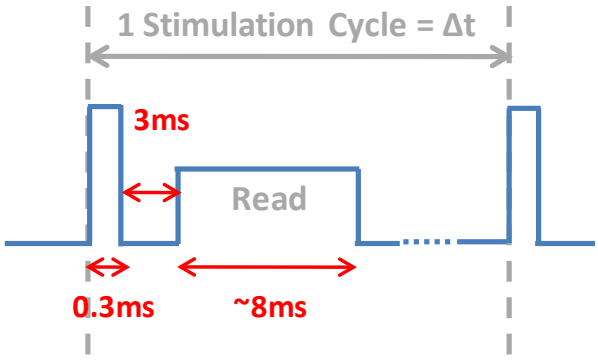


Figure 6.3 Illustration of the waveform used in rate-dependency measurements. Stimulation pulse: +1.3 V, 0.3 ms, read pulse: +0.5 V, 8 ms, and delay time between the stimulation and read pulse: 3 ms.

Figure 6.4(a) shows the response of the memristor to different stimulation rates where the memristor conductance (represented by the currents I_1, I_2, ..., I_{10} acquired during read pulses) was recorded immediately after each stimulation pulses. For Δt = 10 sec (orange bars), barely any increase in current is seen, whereas for Δt = 15 ms (red bars), the upward trend is apparent. This effect is more clearly illustrated in Figure 6.4(b), where the net current increase (ΔI) for each pulsing interval is plotted. Here ΔI was

calculated by offsetting the current (I_N, N = 1, 2, ..., 10) by I_1. A clear dependence of the conductance enhancement on the stimulation rate is unambiguously observed, with a high stimulation rate being the most effective and low stimulation rate being the least effective.

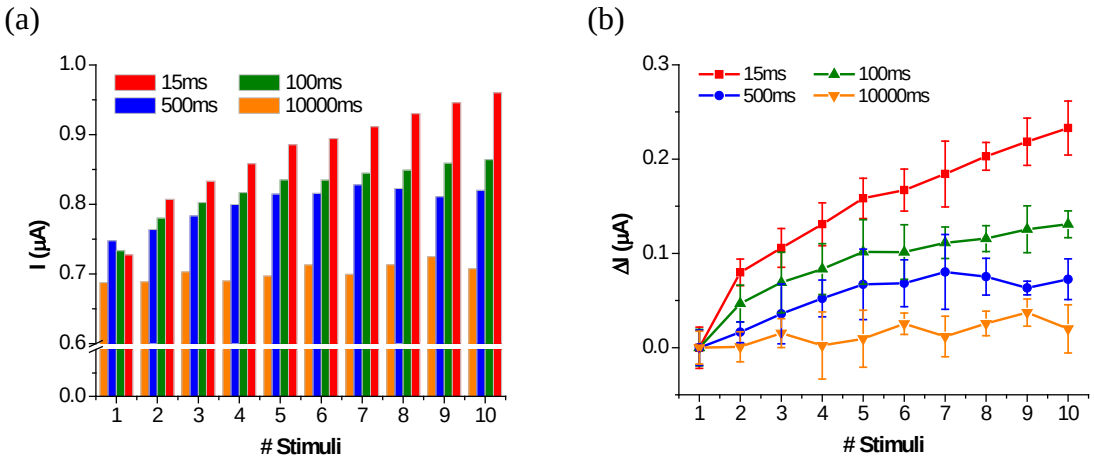


Figure 6.4 Rate-dependent synaptic plasticity of WO_X memristors. (a) Currents read after each stimulus at different stimulation intervals. (b) Currents offset by the first read current. The measurements were repeated five times to minimize fluctuations in data, with solid marks and error bars representing the mean and standard deviation, respectively.

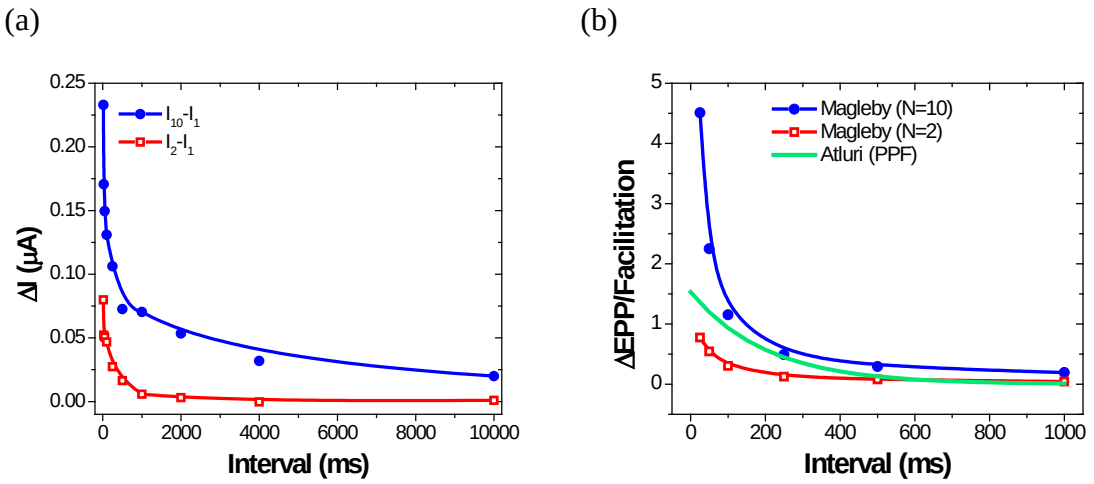


Figure 6.5 PPF and PTP effects observed in memristors and in biological systems. (a) Extracted I_2 - I_1 and I_{10} – I_1 versus stimulation interval. (b) PPF and PTP calculated using the methods and parameters in [4] and [6], respectively. Details on data generation of curves in (b) are provided in the Appendix.

In order to form analogy between the conductance enhancement phenomenon and biological effects, (I_2 - I_1) and (I_{10} - I_1) for each interval condition are plotted in Figure 6.5(a), which correspond to the conditions used for PPF and PTP used in biological studies, respectively. For convenience, rate-dependent effects observed in biological systems are provided in Figure 6.5(b). Indeed, Figure 6.5(a)-(b) show a similar trend and verify the effectiveness of using memristors to emulate the rate-dependent learning rules.

The reason for the rate dependence in biological systems can be briefly explained as the following. A stimulus (action potential) at the presynaptic neuron permits calcium (Ca^{2+}) influx that initiates the release of neurotransmitters, thereby temporarily enhancing synaptic transmission. Once the stimulus is terminated, it takes a finite time for the residual Ca^{2+} to decay to its equilibrium level. Hence, if another identical stimulus succeeds shortly after this stimulus, the response of the synapse is enhanced, as in PPF. In the same fashion, if many stimuli follow closely one after another, synaptic transmission progressively grows with increasing number of stimuli, as in PTP. In WO_X memristors, similar competing effects exist between the stimulation pulses and the diffusion of V_{OX}, resulting in similar rate-dependent behaviors. These comparisons again verify the feasibility of using memristors to emulate synapses for neuromorphic systems.

6.3 Timing-Dependent Synaptic Plasticity

Not only does the spiking interval matter, the spiking order is also found to be crucial in biological systems. A synapse receives signals from both the pre-synaptic and the post-synaptic neurons, and depending on the time of arrival of signals from the two neurons, both the *sign* and *magnitude* of synaptic weight change can differ. This learning

rule is the famous spike-timing dependent plasticity (STDP) [7-8]. In short, if the pre-synaptic neuron spikes before (after) the post-synaptic neuron does, the synaptic weight is strengthened (weakened), i.e. the synapse is being potentiated (depressed); additionally, the shorter the spike timing difference, the larger the synaptic weight change.

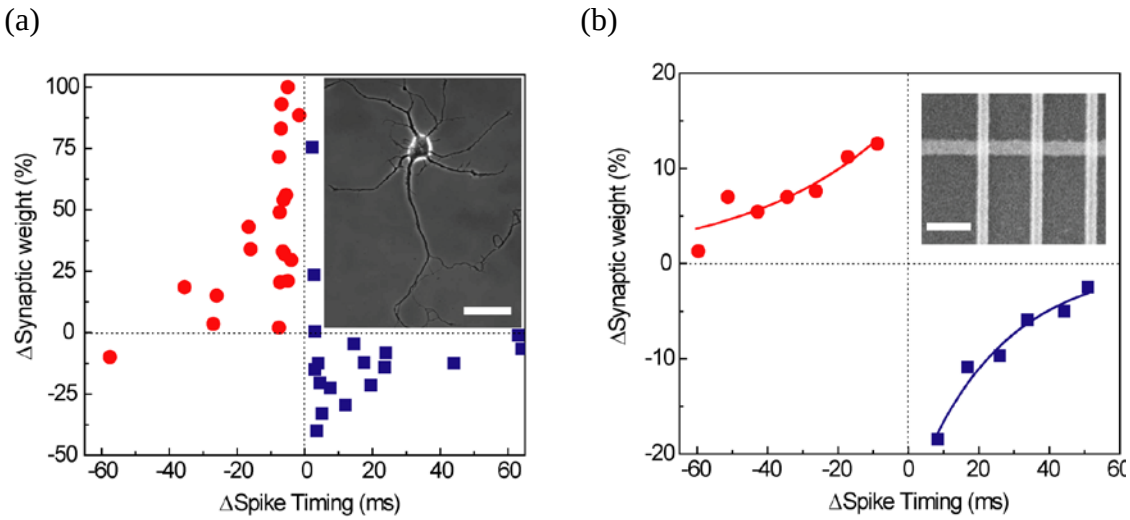


Figure 6.6 STDP observed in a biological synapse and demonstrated in a memristor synapse. (a) The measured change in excitatory postsynaptic current (EPSC) of rat hippocampal neurons after repetitive correlated spiking (60 pulses at 1 Hz) *vs.* relative spike timing. Inset: a phase contrast image of a hippocampal neuron. Scale bar: 50 μm. (b) The measured change of the memristor synaptic weight vs. the relative timing of the neuron spikes. The synaptic weight change was normalized to the maximum synaptic weight. Inset: an SEM image of a Si/Ag co-sputtered memristor crossbar array. Scale bar: 300 nm [9].

In our previous work, STDP was demonstrated for the first time in a Si/Ag co-sputtered memristive device using CMOS circuits as firing neurons [9]. The implementation employed a time division multiplexing (TDM) method with globally synchronized time frames to convert the timing information into pulse widths [10]. Figure 6.6 shows the comparison between the STDP behaviors observed in a biological synapse and in a memristor. Others have subsequently demonstrated STDP in various

devices and material systems [11-15] with different timing techniques but all involves converting timing information into pulse height/width information using biologically-unrealistic programming pulse designs.

In essence, as long as the memristors respond to pulse width and/or pulse height variations in a predictable fashion, realization of STDP comes straightforward with the aid of external neural circuits to provide the necessary pulse design. The TDM method is more complicated as it requires multiplexing units (with the advantage that the pulses having the identical height which is easier to implement in digital circuits); a simpler approach is to design neuron spiking waveforms that automatically translate timing information into pulse width/height profiles. As shown in Figure 6.7, depending on the time difference between the pre- and post-synaptic spikes, the overlapping duration/amplitude (shaded area) would vary accordingly.

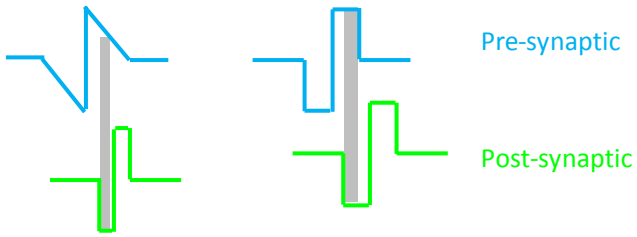


Figure 6.7 Examples of spiking waveforms for translating STDP timing information into overlapping pulse width and/or height (shaded area) [16].

Figure 6.8 shows the nonlinear memristor response to pulse height which reflects the non-linear ionic drift (*sinh* function) in Equation 4.10. Similarly, Figure 6.9 shows the memristor response to pulse width at a fixed pulse height. Achieving STDP using the overlapping profiles shown in Figure 6.7 or the TDM approach is straightforward based on these properties.

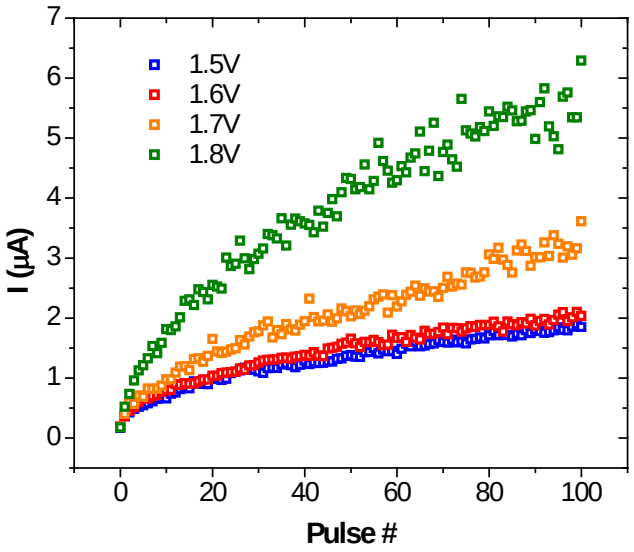


Figure 6.8 Memristor response to stimulation pulse height. Stimulation pulses had 500 μs duration with different amplitudes. Data show the read current during read pulses of 0.3 V, 6 ms applied after each stimulation pulse.

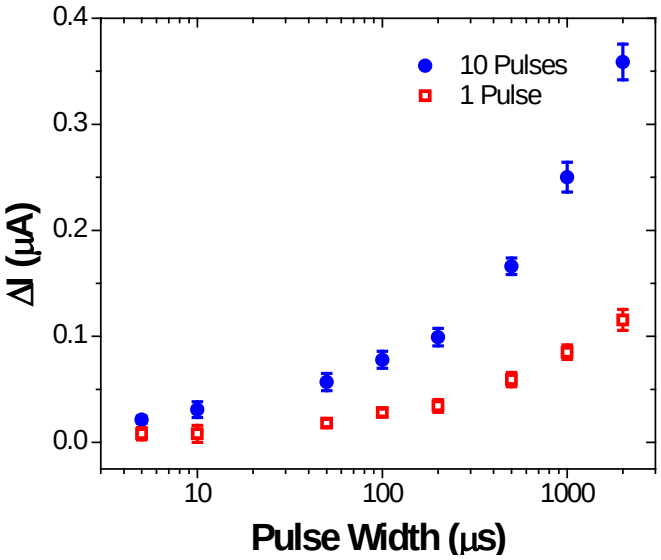


Figure 6.9 Memristor response to stimulation pulse width. Stimulation pulses had 1.3 V amplitude with different durations (5, 10, 50, 100, 200, 500, 1000, and 2000 μs). Data show the read current during read pulses of 0.45 V, 6 ms applied after each stimulation pulse. Each box chart contains statistics from 10 measurements.

6.4 Sliding Threshold Effect

The most generalized synaptic learning rule thus far is the BCM (Bienenstock-Cooper-Munro) theory, which accurately predicted developmental synapticity and neuron selectivity observed experimentally [17]. The theory defined a frequency threshold (θ_m) separating LTP and LTD that is determined by the correlation between pre- and post-synaptic activities. When the stimulation frequency is above θ_m, LTP occurs, and below which, LTD occurs. A consequence of the BCM theory is that rather than being a fixed value, θ_m is a function of stimulation history, thus termed *sliding threshold* or *modification threshold* [18]. Figure 6.10 illustrates the concept of the BCM theory. As the synapse is subjected to high frequency stimulation (HFS), θ_m moves toward the right thus raising the threshold to induce further LTP (light blue curve); meanwhile, as the synapse is subjected to low frequency stimulation (LFS), θ_m moves toward the left making it much easier to induce LTP than LTD (light green curve). This sliding threshold behavior is important for synaptic stability as the synaptic plasticity is always kept within the dynamic range instead of being saturated at boundaries.

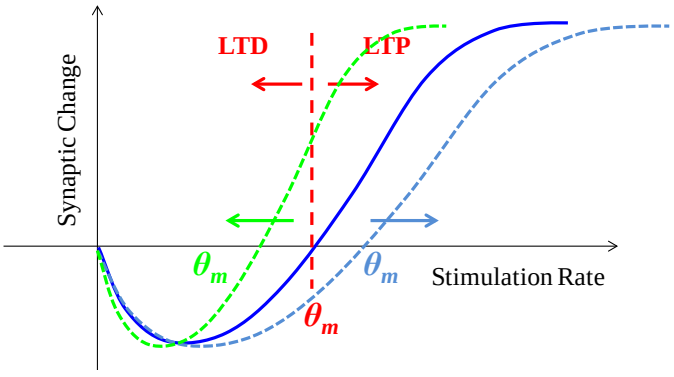


Figure 6.10 Illustration of the BCM theory and the sliding threshold effect.

In [17], the authors formulated the expression for θ_m in terms of the product of the input at the pre-synaptic terminal and the history of activities (i.e. the average firing rate) at the post-synaptic terminal, where the post-synaptic firing rate is a power-function of the summed inputs from each pre-synaptic terminal scaled by their corresponding weights. A simplification of such rule is to assume that the post-synaptic neuron is only connected to one pre-synaptic neuron through the synapse of interest, such that the post-synaptic activity only reflects the activity at this single pre-synaptic neuron. Then θ_m is solely determined by the stimulation history at the pre-synaptic terminal. Experimentally, to demonstrate the sliding threshold effect in memristors, identical stimulating pulses were applied to the pre-synaptic terminal at varying frequencies, as shown in Figure 6.11 (black ticks). It is clear that during 40Hz stimulation (HFS) the current continually increased and during 1Hz stimulation (LFS) the current continually decreased. However, whether it potentiates or depresses is not only determined by the absolute value of the stimulation frequency, instead, the sign and efficiency of the synaptic weight change is also affected by the previous stimulation history. For example, when 10Hz stimulation followed 40Hz stimulation, a downward trend was observed (orange arrow), whereas when identical 10Hz stimulation followed 1Hz stimulation, an upward trend was observed (green arrow). This phenomenon indicates that the proceeding 40Hz stimulation caused θ_m to right-shift, thus making it easier to induce LTD than LTP; while the proceeding 1Hz stimulation caused θ_m to left-shift, thus making it easier to induce LTP than LTD. In addition, whenever the stimulation rate changed, the next few stimuli were always most affected. Therefore, the threshold for inducting LTP and LTD had shifted over time depending on the stimulation history.

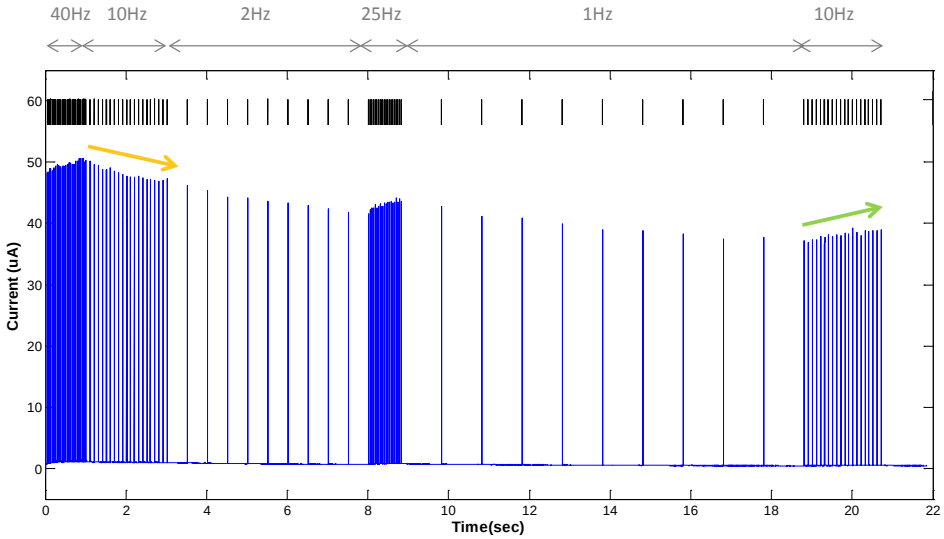


Figure 6.11 Frequency-varying measurements suggest a sliding threshold behavior in WO_X memristors. All pulses were +1.2 V, 1 ms. Device size: 200 nm x 200 nm.

The BCM theory is considered the most universal rule as it incorporates and reconciles both the rate- and timing-dependent synaptic plasticity [19], so it is capable of predicting random neuron-synapse networks with complicated firing patterns. The most significant implication of the BCM theory is that the synaptic plasticity is a dynamic process involving the whole stimulation history. Neuroscientists have proposed a unified model based on the dynamics of Ca^{2+} that successfully related all the details of the BCM theory [20-23]. Research on emulating the entire BCM theory employing the ion dynamics in the memristors is currently underway. However, it should be pointed out that the sliding threshold effect demonstrated here already is one of the crucial aspects of the comprehensive BCM theory. On a side note, the sliding threshold behavior shares a matching concept with "metaplasticity," which literally means the plasticity of synaptic plasticity [24].

6.5 Homosynapticity vs. Heterosynapticity

Synaptic learning can be homosynaptic or heterosynaptic [25-26]. The former refers to situations where synaptic modulation takes place at the same synapse which was being triggered (Figure 6.12(a)), and the latter refers to synaptic modulation that is influenced by a third neuron, , a modulatory neuron (Figure 6.12(b)). In normal two-terminal WO_X memristors that have been discussed so far, plasticity was homosynaptic. To realize heterosynapticity, the device geometry was modified to include a third terminal to act as the modulatory neuron. As shown in Figure 6.13, one extra electrode "MOD" is added to mimic the modulatory neuron, so that through V_{MOD} the plasticity between PRE and POST can be modulated.

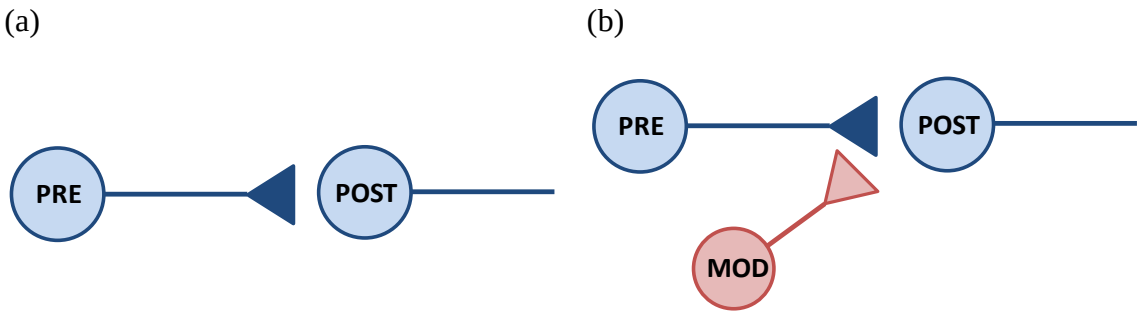


Figure 6.12 Homosynapticity vs. heterosynapticity. (a) Modulation depends on PRE and POST activities. (b) Modulation also affected by the MOD activities.

(a) (b)

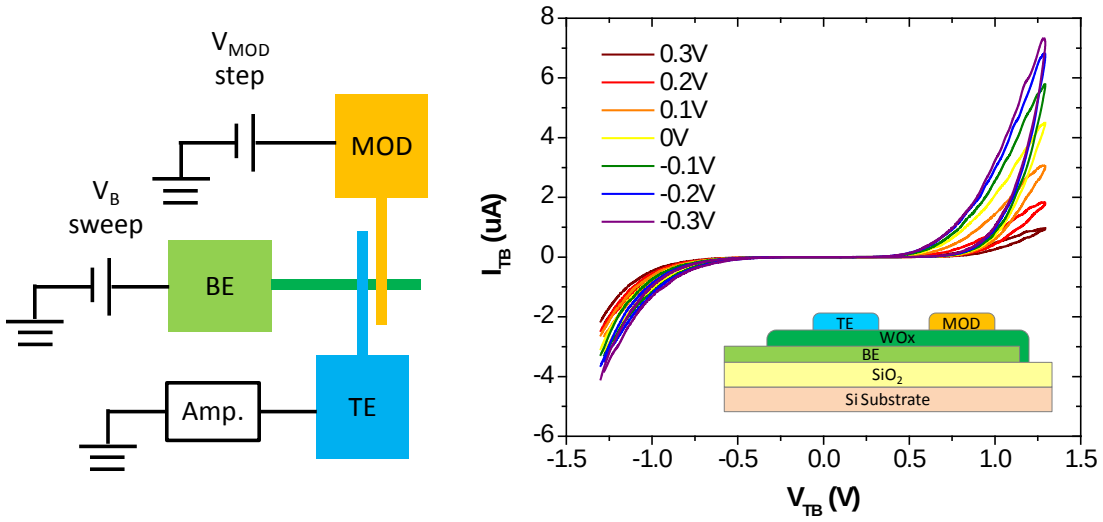


Figure 6.13 Heterosynaptic plasticity in a three-terminal WO_X memristor. (a) Device structure of a three-terminal WO_X memristor and testing setup (not drawn to scale). (b) Modulated switching at the TE-BE junction as a function of V_{MOD}. Cross-section schematic of the device structure is provided as the inset (not drawn to scale). All three electrodes are ~ 170 nm wide, and the distance between TE and G electrodes is ~ 850 nm.

Since RS is controlled by electric field between the TE and BE, RS can be controlled by V_{MOD} by controlling the potential on BE, and this effect enhances as the MOD and TE electrodes get closer. Figure 6.13 shows the preliminary data from a three-terminal memristor where the effects of the MOD electrode on TE-BE switching are clearly observed. Unlike normal memory array operation, here crosstalk between electrodes is necessary. Further studies are ongoing to quantitatively analyze the results and to achieve better correspondence with biology.

6.6 Conclusion

In this chapter, several well-accepted synaptic learning rules observed in biological systems were introduced, including rate-dependent plasticity, timing-dependent plasticity, the BCM theory, as well as homosynaptic and heterosynaptic learning rules. The very important rate-dependent plasticity has been demonstrated experimentally with WO_X memristors along with the gradients for demonstrating timing-dependent plasticity. The WO_X memristors also exhibit the sliding threshold behavior, which is a signature of synapses responding dynamically to neuron firings according to past activities. Moreover, heterosynapticity have been implemented by adding a third electrode to act as a modulatory neuron. The ability of synapses to encode neuron spiking patterns and change weights accordingly enables effective learning in neural networks, therefore the experimental demonstrations provided in this chapter are extremely important in order to build neural networks that can potentially function similarly as their bio-counterparts.

Appendix

Generation of PPF and PTP curves

The green curve (PPF) in Figure 6.5(b) was calculated from an exponential decay function of PPF between the granule/Purkinje cell synapse [4], where the amplitude and the time constant were 153% and 203 ms, respectively. This curve describes how much the response to the second pulse is facilitated comparing to the first pulse with respect to the pulse-to-pulse interval. The blue and red marks in Figure 6.5(b) were calculated using

the model in [6] using scheme II, and all the parameters used in the calculation were obtained from the legend of Figure 4 in the same reference.

Reference

[1] W. M. Cowan, Ed. *Synapses*; Johns Hopkins University Press, 2001.

[2] P. J. Sjöström, G. G. Turrigiano, S. B. Nelson, "Rate, Timing, and Cooperativity Jointly Determine Cortical Synaptic Plasticity," *Neuron* **32**, 1149, 2001.

[3] K. L. Magleby, "The Effect of Repetitive Stimulation on Facilitation of Transmitter Release at the Frog Neuromuscular Junction," *J. Physiol.* **234**, 327, 1973.

[4] P. P. Atluri, W. G. Regehr, "Determinants of the Time Course of Facilitation at the Granule Cell to Purkinje Cell Synapse," *J. Neurosci.* **16**, 5661, 1996.

[5] T. Chang, S. H. Jo, W. Lu, "Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor," *ACS Nano* **5**, 7669, 2011.

[6] K. L. Magleby, "A Quantitative Description of Stimulation-Induced Changes in Transmitter Release at the Frog Neuromuscular Junction," *J. Gen. Physiol.* **80**, 613, 1982.

[7] G. Q. Bi, M. M. Poo, "Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type," *J. Neurosci.* **18**, 10464, 1998.

[8] R. C. Froemke, Y. Dan, "Spike-Timing-Dependent Synaptic Modification Induced by Natural Spike Trains," *Nature* **416**, 433, 2002.

[9] S. H. Jo, T. Chang, I. Ebong, B. Bhavitavya, P. Mazumder, W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems, " *Nano Lett.* **10**, 1297, 2010.

[10] G. S. Snider, "Spike-Timing-Dependent Learning in Memristive Nanodevices," *IEEE/ACM NANOARCH*, Anaheim, CA, June 2008.

[11] G. Indiveri, E. Chicca, R. Douglas, "A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses with Spike-Timing Dependent Plasticity," *IEEE Trans. Neural Netw.* **17**, 211, 2006.

[12] G. Rachmuth, H. Z. Shouval, M. F. Bear, C. S. Poon, "A Biophysically-Based Neuromorphic Model of Spike Rate- and Timing-Dependent Plasticity," *Proc. Natl. Acad. Sci. USA* **108**, E1266, 2011.

[13] D. Kuzum, R. G. D. Jeyasingh, B. Lee, H. -S. P. Wong, "Nanoelectronic Programmable Synapses Based on Phase Change Materials for Brain-Inspired Computing," *Nano Lett.* **12**, 2179, 2012.

[14] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, H. Hwang, "Analog Memory and Spike-Timing Dependent Plasticity Characteristics of a Nanoscale Titanium Oxide Bilayer Resistive Switching Device," *Nanotechnology* **22**, 254023, 2011.

[15] P. Krzysteczko, J. Munchenberger, M. Schafers, G. Reiss, A. Thomas, "The Memristive Magnetic Tunnel Junction as a Nanoscopic Synapse-Neuron System," *Adv. Mater.* **24**, 762, 2012.

[16] Y. V. Pershin, M. Di Ventra, "Neuromorphic, Digital, and Quantum Computation with Memory Circuit Element," *Proc. IEEE* **99**, 1, 2011.

[17] E. L. Bienenstock, L. N. Cooper, P. W. Munro, "Theory for the Development of Neuron Selectivity: Orientation Specificity and Binocular Interaction in Visual Cortex," *J. Neurosci.* **2**, 32, 1982.

[18] M. F. Bear, "Mechanism for a Sliding Synaptic Modification Threshold," *Neuron* **15**, 1, 1995.

[19] E. M. Izhikevich, N. S. Desai, "Relating STDP to BCM," *Neural Comput.* **15**, 1511, 2003.

[20] H. Z. Shouval, M. F. Bear, L. N. Cooper, "A Unified Model of NMDA Receptor-Dependent Bidirectional Synaptic Plasticity," *Proc. Natl. Acad. Sci. USA* **99**, 10831, 2002.

[21] H. Z. Shouval, G. C. Castellani, B. S. Blais, L. C. Yeung, L. N. Cooper, "Converging Evidence for a Simplified Biophysical Model of Synaptic Plasticity," *Biol. Cybern.* **87**, 383, 2002.

[22] R. S. Zucker, W. G. Regehr, "Short-Term Synaptic Plasticity," *Annu. Rev. Physiol.* **64**, 355, 2002.

[23] M. Graupner, N. Brunel, "Calcium-Based Plasticity Model Explains Sensitivity of Synaptic Changes to Spike Pattern, Rate, and Dendritic Location," *Proc. Natl. Acad. Sci. USA* **109**, 3991, 2012.

[24] W. C. Abraham, M. F. Bear, "Metaplasticity: the Plasticity of Synaptic Plasticity," *Trends Neurosci.* **19**, 126, 1996.

[25] C. H. Bailey, M. Giustetto, Y. Y. Huang, R. D. Hawkins, E. R. Kandel, "Is Heterosynaptic Modulation Essential for Stabilizing Hebbian Plasticity and Memory?" *Nat. Rev. Neurosci.* **1**, 11, 2000.

[26] G. A. Phares, J. H. Byrne, "Heterosynaptic Modulation of Synaptic Efficacy," Encyclopedia of Life Sciences, 2006.

Chapter 7
Integration of Memristor Arrays with CMOS Circuits

7.1 Introduction

The advantage of memristors and the crossbar structure associated with them will only be useful if each memristor in an array can be individually accessed and given desired operations. Regarding the operation of the crossbar arrays, the cells need to first be addressed and then be applied with a program, erase, or read voltage. When a read voltage is applied, the current is sensed by a resistor or an amplifier. However, an inherent problem with the crossbar structure is the sneak path problem in which the half-selected cells carry a large reverse current resulting in a faulty read [1]. Several possible solution to this problem have been proposed, for example, using a transistor or a diode as the selecting device (1T1R or 1D1R) [2-3], adding a soft-breakdown or volatile-switching device in series [4-5], designing a memory device with intrinsic rectification [6-7], or connecting two RRAM (memristor) devices anti-serially [8]. To date no clear winder has been established on how to solve the sneak path problem yet.

The peripheral circuitry is by and large CMOS-based due to its reliability and versatility, so there are several factors to consider. First, the memristor has to be compatible with CMOS circuits both fabrication- and the operation-wise, i.e. not only the fabrication needs to follow the standard processes, the operating voltage, current level,

power dissipation, and switching speed should also be CMOS-comparable. Second, since memristors are analog devices, the sensing circuitry has to be sensitive enough to distinguish different conducting states. Third, the short retention within WO_X memristors complicates the program/erase/read scheme as the state may have decayed after individually addressing each cell in an array. Lastly, at the system level, the CMOS neuron circuits require careful design in order to work collaboratively with memristor synapses forming functional neural networks. The fabrication of memristor arrays on CMOS chips and the operation of such arrays will be covered in this chapter.

7.2 Fabrication of Memristor Arrays on 180 nm CMOS Chips

The first attempt was to fabricate memristor arrays on top of CMOS chips which had multiplexer/de-multiplexer (MUX/DeMUX) circuits designed by our collaborators at HRL Laboratories and taped-out with IBM's 7RF process. Recently, the integration of 1.6 Kb digital memristor devices arrays with these CMOS chips have been reported by our group [9]. For analog devices, a similar approach was used. We began from smaller scales, namely, 10-by-10 arrays. At the 180 nm generation node W inter-level connections were used, so the top surface of these chips was W vias and SiN passivation planarized by CMP. As can be seen from Figure 7.1(a), the bright dots are the W vias embedded in the SiN passivation layer.

Two additional fabrication steps were required for integrated devices compared to standalone ones. First, during the oxidation step to form WO_X, the top surface of the W vias would be oxidized at the same time, inhibiting good metal contacts to the buried interconnects. To avoid this problem, certain vias were covered with inert metal to

guarantee good metal contacts. Second, thinner PMMA was used to achieve smaller pitch size (that is, higher density) so spin-on-glass (SOG) was introduced to eliminate step jumps. Figure 7.1(b)-(d) show the completed 200 nm pitch 10 x 10 analog memristor array built on the 180 nm CMOS chip. The detailed fabrication steps are provided in Table 7.1.

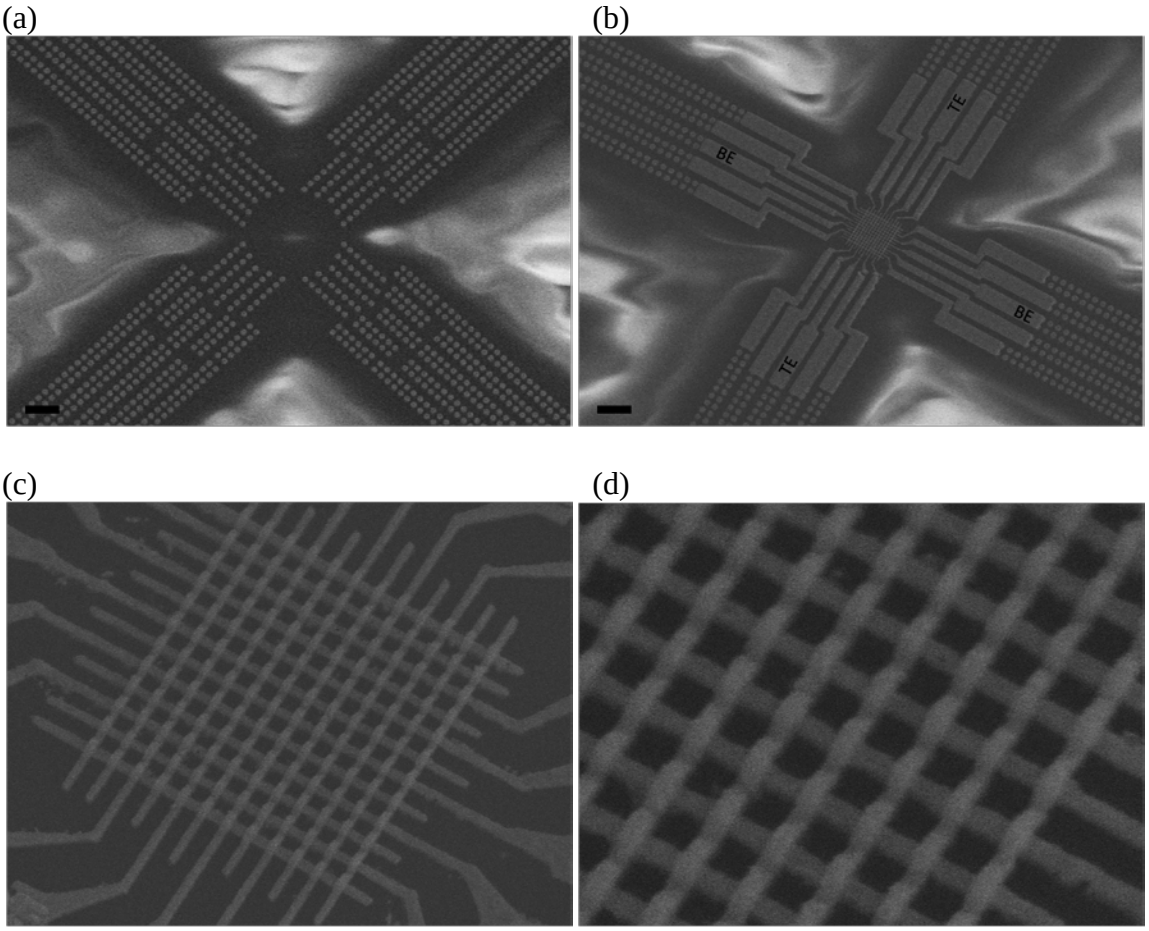


Figure 7.1 SEM images of the integrated 180 nm CMOS chip. (a) CMOS chip terminated at planarized W vias and SiN passivation. (b) A 10 x 10 memristor array fabricated on the same chip. Scale bars: 2 μm; 25° tilt view. (c) and (d) are zoom-in views of (b). The pitch is 200 nm and each electrode is ~80 nm wide.

Table 7.1 Process flow for WO_X memristor array integration on 7RF CMOS chips.

Step	Process	Comment
	Photoresist removal	Removal by Remover PG (MicroChem)
	W deposition	60 nm W by DC Sputtering at room temperature (Kurt. J. Lesker LAB18)
	EBL for BE & via connection	EBL (Raith 150) using PMMA A2 resist, MIBK:IPA = 1:3 developer (MicroChem)
	Ni evaporation & liftoff	40 nm Ni hard mask by e-beam evaporation (Cooke Evaporator), liftoff in acetone
	W etch	RIE using Cl_2/O_2 gas mixture (LAM 9400)
	Ni removal	Removal of Ni by HCl:DI solution
	Sidewall passivation	PECVD at 200 °C (GSI PECVD)
	SOG planarization	SOG deposition and RIE etchback
	EBL for via covering	EBL using PMMA A4 resist, MIBK:IPA = 1:3 developer
	NiCr/Pd evaporation & liftoff	E-beam evaporation, liftoff in acetone
	Spacer formation	PECVD at 200 °C, RIE etchback
	WO_X formation	Oxidation in O_2 ambient (JetFirst-150 RTP)
	EBL for TE	EBL using PMMA A2 resist, MIBK:IPA = 1:3 developer
	Pd/Au evaporation & liftoff	Ee-beam evaporation, liftoff in acetone
	WO_X etch	RIE using SF_6/C_4F_8 gas mixture

Primitive testing showed that addressing each cell and providing voltage biases were feasible utilizing our custom-built MATLAB program and hardware interface. Figure 7.2(a) is an example of a DC sweep from 0 V to 2 V; Figure 7.2(b) shows the

continuous increase of memristor conductance until reaching the CC. The programming voltage here is slightly higher than that of standalone devices, possibly due to the additional resistances contributed from via contacts (via diameter ~ 360 nm) and the underlying CMOS circuits.

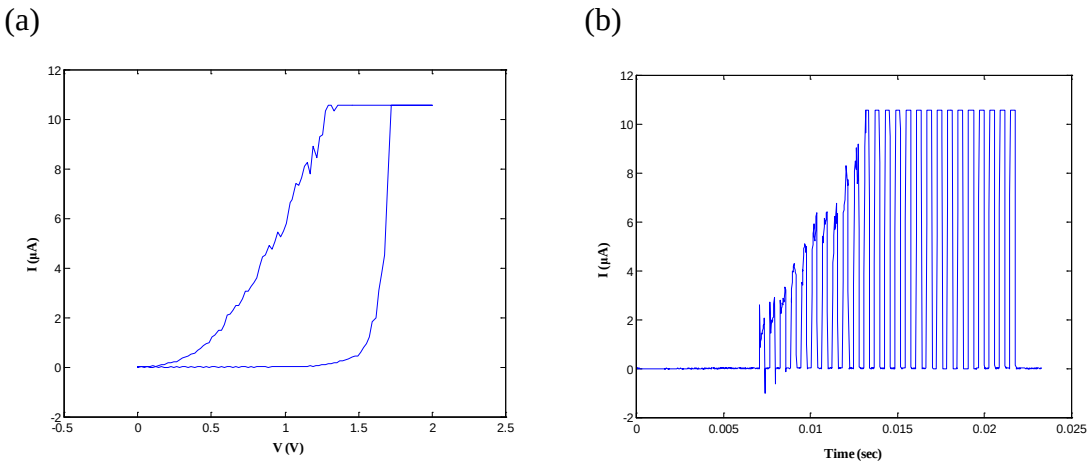


Figure 7.2 Testing of integrated 10 x 10 memristor arrays on 7RF CMOS chips. (a) DC sweep (0 V → +2 V, 2 V/sec). (b) Pulse operation (+1.8 V, 300 μs, 25 pulses).

7.3 Fabrication of Memristor Arrays on 90 nm CMOS Chips

Similar to the 180 nm chips, the 90 nm chips had CMOS circuits taped-out with IBM's 9RF process by our collaborators at HRL Laboratories. However, integration with 90 nm chips is not as straightforward as integration with 180 nm chips since at the 90 nm generation node Cu is used for inter-level connections instead of W. Cu has better conductivity but oxidizes in air very easily. In addition, Cu is not as chemically stable as W. Thereof a SiN passivation layer was deposited on top of Cu by the foundry to prevent its exposure to air during shipping and handling. Prior to memristor array fabrication, the

vias need to be etched and filled with inert metals to make electrical contact to the buried Cu wires and extend these interconnects to the surface.

7.3.1 Via-Etching Processes

SiN etching is a crucial step in the Cu dual-damascene process [10]. When the SiN passivation layer is etched through, Cu surface becomes contaminated with post-etch residues of polymers (resist masking layer), by-product of SiN etch, and/or redeposited materials [11]. One mistake that should always be avoided is the exposure of Cu surface to etched polymers as polymer residues are stripped using O_2 plasma which potentially oxidizes Cu. This can be solved by using other dielectric layers as a sacrificial mask for etching SiN [12]. In fact, multiple sacrificial dielectric layers are used in commercial manufacturing processes, including SiC, undoped silicate glass (USG), SiOCH low-κ films, and bottom anti-reflective coatings (BARC). As shown in Figure 7.3(a), a blank film of PECVD SiO_2 was first deposited. The SiO_2 sacrificial layer was etched using patterned PMMA A4 resist as mask (Figure 7.3(b)) and PMMA was afterward removed by O_2 plasma by RIE (Figure 7.3(c)). By doing this, Cu was protected during etching and removing PMMA, which significantly helped in keeping Cu etched surface clean and un-oxidized. Finally, the SiN was etched using SiO_2 as mask with a gas mixture that etches SiO_2 and SiN at equal rates (Figure 7.3(d)).

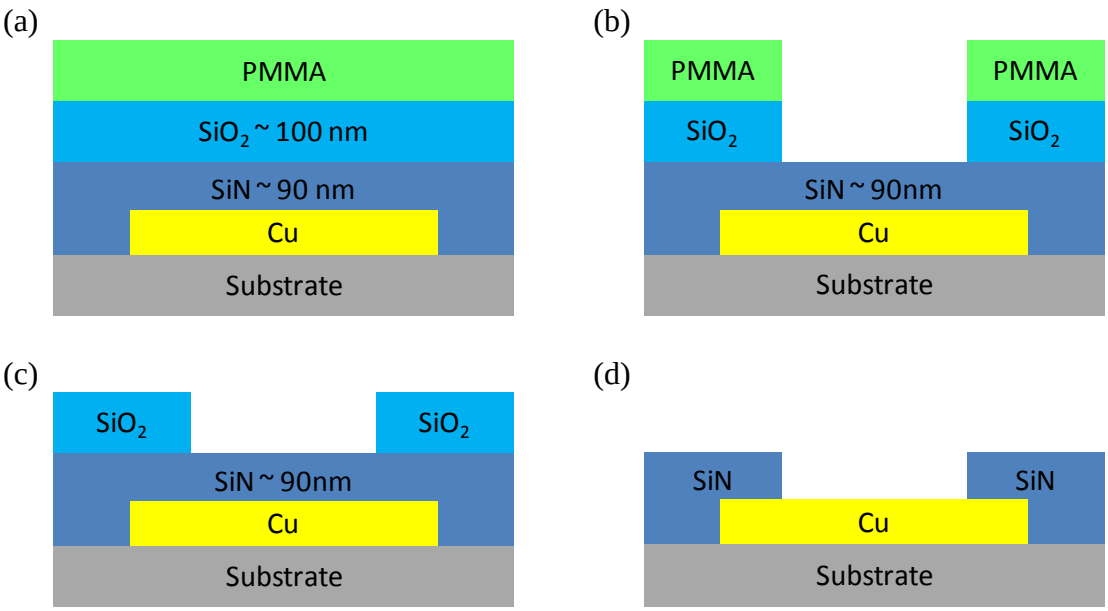


Figure 7.3 Via-etching process flow. Vias are ~600 nm in diameter.

One more critical step to the via-etching process is post-etch cleaning, which should remove all possible post-etch residues as well as copper oxides (including sub-oxides). Many acids and a few bases have been utilized for this purpose; there are also commercial solutions that even incorporate some sort of inhibitors to prevent Cu from oxidizing [13-14]. Different dilution ratios of hydrofluoric acid (HF), buffered-hydrofluoric acid (BHF), hydrogen chloride (HCl), acetic acid, and citric acid have been tested. We found that the 1:1 diluted citric acid is effective in removing post-etch residues yet not eroding the Cu surface. Figure 7.4(a) and (b) show the SEM images of etched vias before and after acid cleaning treatment, respectively. With proper cleaning, the grainy surface of Cu is clearly observed, as in Figure 7.4(b).

(a) (b)

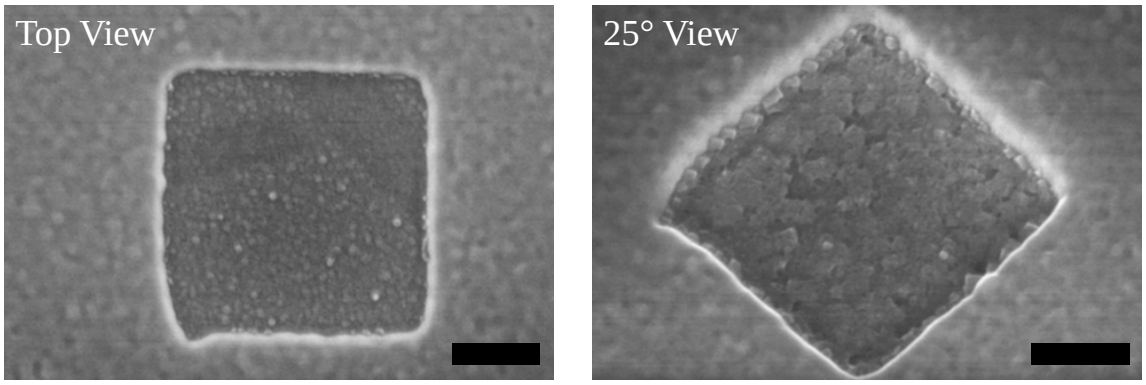


Figure 7.4 SEM images of etched vias. (a) Immediately after SiN etch and (b) immediately after acid cleaning treatment. Scale bars: 250 nm. The sample was cleaned in citric acid : DI water = 1 : 1 solution for 1 min at room temperature.

Characterization of via-etching process is through the via-chain structure [15], which evaluates the summed resistances contributed by different numbers of vias. The process flow for fabricating via-chains is provided in Table 7.2. Figure 7.5 shows an optical microscope image of such structure. Through controlled RIE etching and wet cleaning processes, a contact resistance of tens of Ohms per via (~650 nm x 650 nm) was obtained, with a 50-via via-chain being the longest fabricated.

Table 7.2 Process flow for via-chain fabrication.

Step	Process	Comment
	Photoresist removal	Removal by Remover PG (MicroChem)
	SiO_2 deposition	100 nm SiO_2 by PECVD at 200 °C (GSI PECVD)
	EBL for via holes	EBL (Raith 150) using PMMA A4 resist, MIBK:IPA = 1:3 developer (MicroChem)
	SiO_2 etch	RIE with PMMA mask (LAM 9400)
	PMMA removal	RIE using O_2 plasma, then acetone bath
	SiN etch	RIE using SF_6/C_4F_8 gas mixture with SiO_2 as mask, but etched SiN and SiO_2 equally
	Post-etch cleaning	Citric acid : DI water = 1 : 1, then DI water rinse
	EBL for via-chains	EBL using PMMA A6 resist, MIBK:IPA = 1:3 developer
	NiCr/Au evaporation & liftoff	10 nm NiCr / 120 nm Au by e-beam evaporation, liftoff in acetone

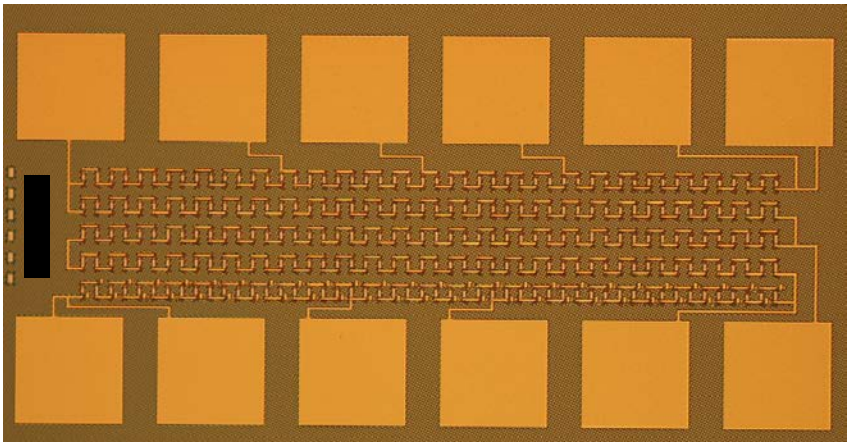


Figure 7.5 Microscope image of a via-chain structure. The horizontal bars are on top of SiN and the vertical bars are Cu underneath SiN. Scale bar: 80 um.

7.3.2 Via-Filling and Planarization Processes

The via-chain structures tested in the previous section was fabricated on test chips which had no built-in CMOS circuitry so there was no following memristor fabrication processes. When memristor fabrication is considered, minor adjustments and optimization are required. Starting from the electrical contacts made by Ti/Au stack, planarization should first be completed to construct a flat surface for memristor fabrication. SOG is the best available material but it also has several drawbacks. First, SOG has a tendency to react with certain metal materials (e.g. W) so a thin protective PECVD SiO_2 layer is always inserted to isolate SOG from metal structures. But the different etch rates between PECVD SiO_2 and SOG causes uncertainty during SOG etch-back and degrades the planarization profile. Second, SOG agglomerates more on large structures (e.g. pads) than on small structures (e.g. memristors' electrodes). This area-dependency of SOG thickness makes it difficult to construct a completely flat surface as both large and small patterns are present. Third, SOG is incapable of planarizing taller structures. The SOG curing recipe used results in a 300 nm thick film, and any structure taller than about half its thickness (~150nm) cannot be completely planarized. With all the issues listed, planarizing tall Ti/Au stacks with SOG is probably not the best idea.

Alternatively, instead of *adding* an SOG layer, one can *subtract* an SOG layer. The via-etching process was identical to that of the via-chain structure, except the sacrificial SiO_2 mask thickness was increased from 100 nm to 150 nm to allow larger etch margin during the planarization process, as shown in Figure 7.6(a). A 120 nm sputtered W layer and a 100 nm PECVD SiO_2 layer were deposited, and about 300 nm of SOG layer was formed, assuming that SOG filled in the via holes and planarized the top

surface (Figure 7.6(b)-(c)). The first global etch step selectively etched SiO_2, leaving the profile like Figure 7.6(d). The second global etch step un-selectively etched SiO_2, W, and SiN at similar rates, finally completing the via-filling process with a profile shown in Figure 7.6(e) and Figure 7.7. The procedures for the via-filling/planarization process are listed in Table 7.3.

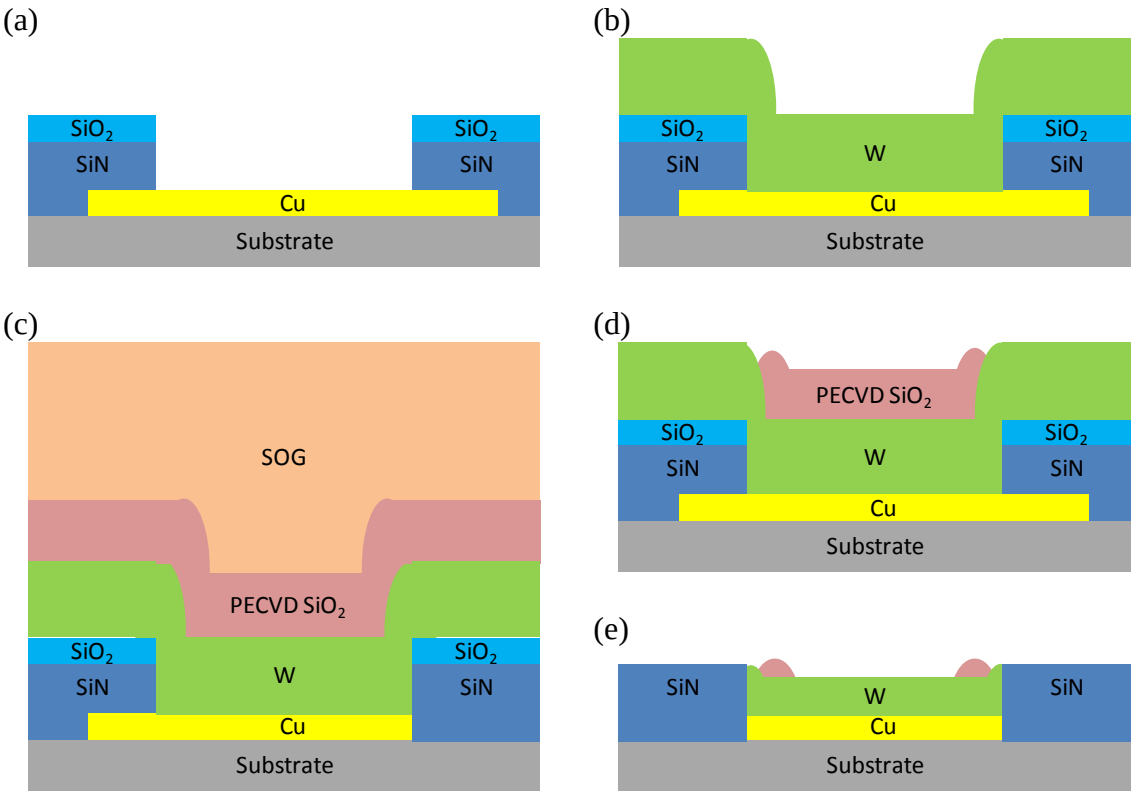


Figure 7.6 Via-filling process flow. Vias are ~500 nm in diameter.

(a) (b)

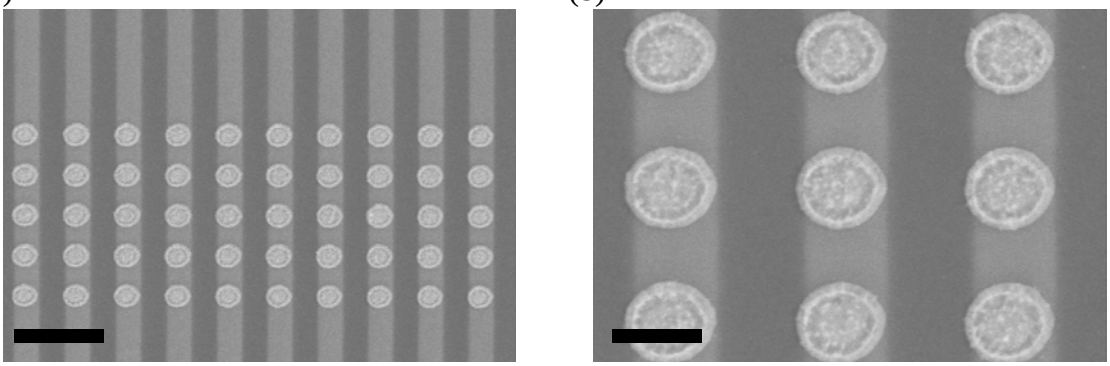


Figure 7.7 SEM images of completed via-filling/planarization process. (a) Top-view. The circles are the W via plugs whereas the vertical lines are the buried Cu wires. Scale bar: 2 μm. (b) Close-up view. Scale bar: 600 nm.

Table 7.3 Process flow for via-filling/planarization fabrication on 9RF CMOS chips.

Step	Process	Comment
	Post-etch cleaning	PRS 1000, then DI water rinse; citric acid : DI water = 1 : 1, then DI water rinse
	Via filling	120 nm W by sputtering (Enerjet), and 100 nm SiO_2 by PECVD at 200 °C, SOG (Filmtronics 700B) spin-coat and cure
	Global etch – I	RIE global etch using chemistry I
	Global etch – II	RIE global etch using chemistry II

This via-filling/planarization process is conceptually similar to the CMP process used in the silicon VLSI technology, except that no mechanical role was introduced here. The process extended the vulnerable Cu lines with inert W plugs allowing easy storage of ongoing samples as well as easy processing in subsequent steps. More importantly, Cu surface was never exposed to resists during any of the lithography, filling, and etching processes, which prevents any residues from being trapped between films, later causing alloying or degassing problems.

7.3.3 Memristor Arrays on 90nm CMOS chips

Once the vias are etched, filled, and planarized, fabrication of memristor arrays on 9RF chips are basically the same as the 7RF processes. Figure 7.8(a) and (b) show the completed 10 x 10 analog memristor arrays built on the 90 nm CMOS chip. There are two different CMOS designs supported so the underlying Cu lines have different layouts. For both chips, the original pitch is designed at 1.12 μm, but there are no constraints to making smaller pitched arrays. The pulse operation of a typical memristor device is shown in Figure 7.9; as in the integrated 7RF devices, the programming voltage is slightly higher than that of standalone devices.

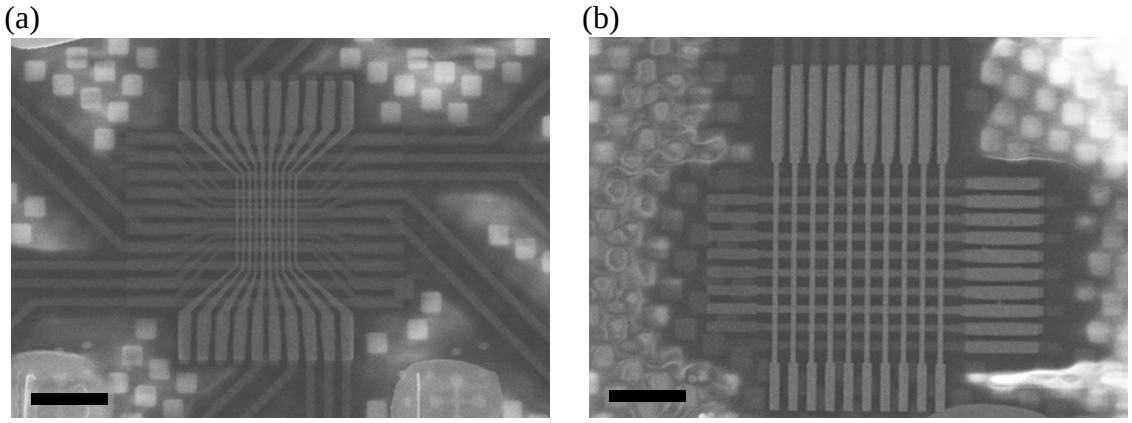


Figure 7.8 SEM images of 10 x 10 memristor arrays integrated on 90 nm CMOS chips. (a) On a direct access chip with 200 nm pitch. (b) On a chip with MUX/DeMUX circuits with 1.12 μm pitch. Scale bars: 5 μm.

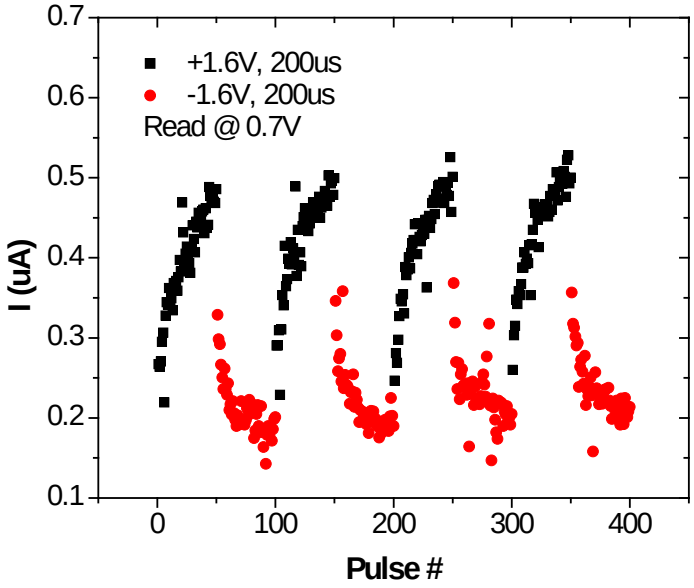


Figure 7.9 Pulse testing of integrated 10 x 10 memristor arrays on 9RF CMOS direct access chips. The data were acquired from a device in an array on the direct access chip. Device size: 400 nm x 400 nm.

7.4 CMOS Circuitry

The 180 nm CMOS MUX/DeMUX circuitry for 10 x 10 memristor arrays is shown in Figure 7.10 and Figure 7.11. On each side (N/W/S/E) are eight contact pads, one of which is for Vdd/GND, two of which are connected to two independent inputs (DATA A and DATA B), and the remaining ones are for providing addresses to the decoder. There are only 5 electrodes connected on each side for a 10 x 10 array, so a 3-bit address is sufficient. When an electrode is selected (a column or a row), it is connected to one voltage input (DATA A) and all the unselected electrodes (other columns and rows) are connected to another voltage input (DATA B). Details on the operation of MUX/DeMUX of these chips are provided in [9].

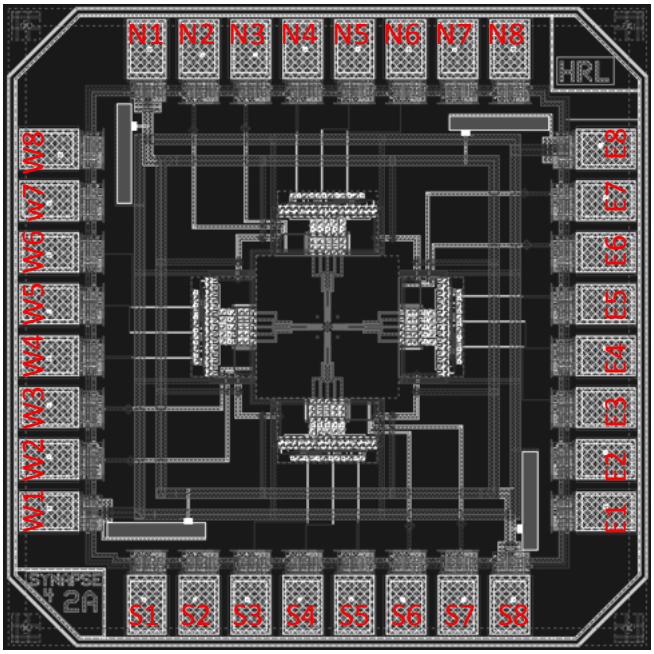


Figure 7.10 Image of MUX/DeMUX circuitry for a 10 x 10 memristor array.

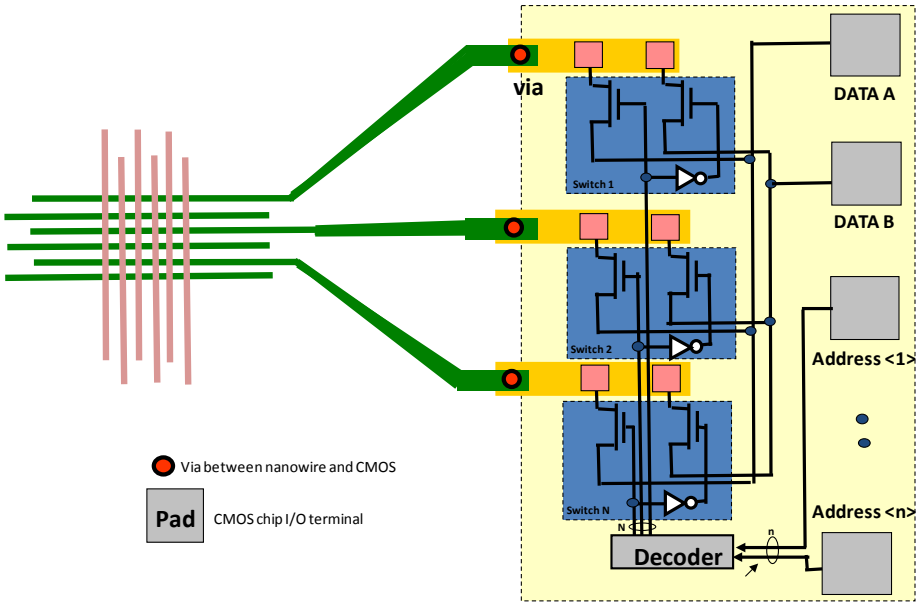


Figure 7.11 Schematic for selecting electrodes in a 10 x 10 memristor array.

The 90 nm CMOS chips support more advanced functions than the 180 nm chips did, for example, built-in write/read circuits for true array operations and CMOS neuron

circuits for large scale neuromorphic computations. In particular, there is one set of chips designed with 100 CMOS neurons interwoven by 10,000 memristor synapses. Thus, integration of memristor arrays with 90 nm chips is not merely to prove process compatibility but, one step further, also to realize functional neuromorphic networks.

7.5 Conclusion

On-chip integration of memristors is an important demonstration for large-scale operations in high-density neural networks. In this chapter, successful integration of 200 nm pitch 10 x 10 memristor arrays on 180 nm CMOS chips is presented, where individual cells in the array behaved normally as if they were standalone. Integration with 90 nm CMOS chips is perplexed by the via-etching process associated with Cu interconnects, but a solution has been proposed and verified. By introducing the sacrificial etch mask and mild post-cleaning processes, good electrical contact to the underlying Cu wires is thus assured. The additional features of the 90 nm chips will soon be explored and utilized.

Reference

[1] M. J. Lee, Y. Park, D. S. Suh, E. H. Lee, S. Seo, D. C. Kim, R. Jung, B. S. Kang, S. E. Ahn, C. B. Lee, D. H. Seo, Y. K. Cha, I. K. Yoo, J. S. Kim, B. H. Park, “Two Series Oxide Resistors Applicable to High Speed and High Density Nonvolatile Memory,” *Adv. Mater.* **19**, 3919, 2007.

[2] S. Kim, H. Y. Jeong, S. K. Kim, S. Y. Choi, K. J. Lee, "Flexible Memristive Memory Array on Plastic Substrates," *Nano Lett.* **11**, 5438, 2011.

[3] B. Cho, T. W. Kim, S. Song, Y. Ji, M. Jo, H. Hwang, G. Y. Jung, T. Lee, “Rewritable Switching of One Diode – One Resistor Nonvolatile Organic Memory Devices,” *Adv. Mater.* **21**, 1, 2009.

[4] R. Meyer, L. Schloss, J. Brewer, R. Lambertson, W. Kinney, J. Sanchez, D. Rinerson, "Oxide Dual-Layer Memory Element for Scalable Non-Volatile Cross-Point Memory Technology," *NVMTS*, Pacific Grove, CA, November 2008.

[5] W. Lee, J. Park, S. Kim, J. Woo, J. Shin, G. Choi, S. Park, D. Lee, E. Cha, B. H. Lee, H. Hwang, "High Current Density and Nonlinearity Combination of Selection Device Based on $TaO_X/TiO_2/TaO_X$ Structure for One Selector – One Resistor Arrays," *ACS Nano* **6**, 8166, 2012.

[6] K. H. Kim, S. H. Jo, S. Gaba, W. Lu, "Nanoscale Resistive Memory with Intrinsic Diode Characteristics and Long Endurance," *Appl. Phys. Lett.* **96**, 053106, 2010.

[7] S. C. Puthentheradam, D. K. Schroder, M. N. Kozicki, "Inherent Diode Isolation in Programmable Metallization Cell Resistive Memory Elements," *Appl. Phys. A – Mater.* **102**, 817, 2011.

[8] E. Linn, R. Rosezin, C. Kügeler, R. Waser, "Complimentary Resistive Switches for Passive Nanocrossbar Memories," *Nat. Mater.* **9**, 403, 2010.

[9] K. H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, W. Lu, "A Functional Hybrid Memristor Crossbar-Array/CMOS System for Data Storage and Neuromorphic Applications," *Nano Lett.* **12**, 389, 2012.

[10] J. D. Plummer, M. D. Deal, P. B. *Griffin, Silicon VLSI Technology: Fundamentals, Practice, and Modeling*; Prentice Hall, 2000.

[11] C. F. Tsang, C. K. Chang, Y. J. Su, H. Y. Li, L. Y. Wong, W. H. Li, M. L. Koh, "Improved Dense Via Yields of Cu/CVD low k Coral^{TM} Dual Damascene Metallization at Post Cu Cap Etch Wet Clean," *Microelectron. Eng.* **75**, 433, 2004.

[12] T. Furusawa, N. Sakuma, D. Ryuzaki, S. Kondo, K. Takeda, S. Machida, R. Yoneyama, K. Hinode, "Direct Resist Removal Process from Copper-Exposed Vias for Low-Parasitic-Capacitance Interconnects," *J. Electrochem. Soc.* **148**, G190, 2001.

[13] C. F. Tsang, C. K. Chang, A. Krishnamoorthy, K. Y. Ee, Y. J. Su, H. Y. Li, W. H. Li, L. Y. Wong, "A Study of Post-Etch Wet Clean on Electrical and Reliability Performance of Cu/low k Interconnections," *Microelectron. Reliab.* **45**, 517, 2005.

[14] D. Louis, A. Beverina, C. Arvet, E. Lajoinie, C. Peyne, D. Holmes, D. Maloney, "Cleaning Status on Low-k Dielectrics in Advanced VLSI Interconnect: Characterisation and Principle Issues," *Microelectron. Eng.* **57-58**, 621, 2001.

[15] C. F. Tsang, V. N. Bliznetsov, Y. J. Su, "Study and Improvement of Electrical Performance of 130 nm Cu/CVD low k SiOCH Interconnect Related to Via Etch Process," *Microelectr. J.* **34**, 1051, 2003.

Chapter 8
Summary and Future Work

8.1 Summary

This thesis work presented the development and advancement of memristors as an emerging technology for neuromorphic applications. WO_X-based memristors were chosen as the model system but much of the knowledge learned can be extended to other memristor system too. Advantages of the WO_X memristor devices include easy fabrication, low-power, high-density, scalability, connectivity, and CMOS-compatibility. In addition, they inherently mimic the functions of biological synapses. Important synaptic functions have been demonstrated, such as rate-dependent plasticity, memory enhancement effect, sliding threshold effect, and heterosynapticity. Integration of WO_X memristors with CMOS technology has been proven feasible, paving the path to actual functionality of memristor arrays.

8.1.1 Development and Understanding of WO_X Memristors

In Chapter 2, the fabrication and characterization of WO_X memristors were presented. The fabrication of WO_X memristors is simple and CMOS-compatible in nature, enabling integration of such devices with CMOS circuits at high density (even 3D stacking). These nanoscale devices consume low power, requiring less than 2 V and

operating at tens of microamperes; they also offer high speed, as resistive switching is based on the movement of the very mobile V_{OX}. High endurance of > 10^8 makes them well suited for neuromorphic applications. Reliable analog switching was demonstrated along with good uniformity across difference devices. Moreover, analog-to-digital switching and current rectification were observed.

In Chapter 3, material characterization techniques and electrical measurements were employed to investigate the film quality and switching characteristics of the WO_X films. XRD and XPS analyses indicated that the RTA-oxidized WO_X films are polycrystalline and near-stoichiometric. Effects of annealing gases, TE materials, and testing atmospheres further shed light on the distribution and concentration of V_{OX} within the WO_X thin film. Joule heating in the conducting filaments was also performed for both analog and digital switching. Counter-intuitively, it seems that V_{OX} is more concentrated near the TE, possibly due to the segregation of V_{OX} at GBs.

Later in Chapter 4, a phenomenological model based on the memristive equations was derived where the internal state variable was assumed to physically represent the width of the conducting region or equivalently the number of parallel conducting paths. The model was successfully implemented in commercial SPICE and numerical softwares. This model is very compact but nonetheless captures most of the switching dynamics and is extremely useful in facilitating circuit-level designs.

8.1.2 Neuromorphic Applications of WO_X Memristors

In Chapter 5, memory retention study was performed and a phenomenon termed memory enhancement effect was characterized. By changing the number of stimuli

applied to the memristor, a clear enhancement in both the conductance and the retention time was seen. Microscopically, parallel conducting paths were created and widened during repetitive stimulation, reducing the probability of completely breaking a conducting path. Thus, the retention time could be increased from less than a second to a minute or so. The memory enhancement effect mimics the transformation of short-term memory to long-term memory through repetition in biological systems by creating more and stronger synaptic connections.

In Chapter 6, different synaptic learning rules were revisited and most importantly implemented in WO_X memristors. Due to the competition between resistive switching and oxygen vacancy diffusion, the higher the stimulation rate, the greater the device conductance change. Stimulation intervals between 15 ms and 10 sec were tested. The rate-dependent synaptic plasticity in WO_X memristors was then compared with biological effects – post-tetanic potentiation and paired-pulse facilitation, which also exhibited a similar competition between stimulation and retention based on the calcium dynamics. Spike-timing dependent plasticity was realized in α-Si/Ag memristors using time division multiplexing method, and could also be achieved in WO_X memristors using designed pulse waveforms. The sliding threshold effect, one of the very important aspects of the BCM rule, is a direct evidence that the synaptic plasticity dynamically depends on the past stimulation history. Demonstration of sliding threshold effect in WO_X memristors suggests that there is an internal feedback mechanism within the film to prevent itself from saturating at boundary states. While homosynaptic plasticity was demonstrated in two-terminal memristors, heterosynaptic plasticity was implemented in three-terminal memristors using a third electrode as the modulatory neuron. It was shown that the MOD

electrode strongly affected the resistive switching effect between the bottom and top electrodes, analogous to the effects of the modulatory neuron on the synaptic weight change between the pre- and post- neurons.

8.1.3 Integration of Memristor-CMOS Hybrid Systems

In Chapter 7, integration of the memristor and CMOS chips at different technology nodes (180 nm and 90 nm) was carried out. Although the fabrication of WO_X memristors is CMOS-compatible, many technical hurdles have to be overcome to ensure proper functioning of both components. In particular, a great amount of work involving etching, cleaning, filling, and planarizing the Cu vias was required in order to accomplish good contact interfaces with Cu interconnects. With a custom-designed peripheral circuits and testing setup, WO_X memristors in 10 x 10 arrays could be addressed and operated individually with success. The switching voltage of integrated memristors was typically 0.1 V – 0.3 V higher than that of standalone ones due to the inevitable series resistance contributed from thin electrodes, via contacts, and CMOS circuits. The integration study is of great importance as it proves that memristors can be successfully integrated on CMOS chips in a BEOL fashion and the integrated chip can work properly as designed. From here, there are great opportunities for different types of applications, especially in reconfigurable logic and neuromorphic systems [1].

8.2 Future Work

8.2.1 In-Depth Material Study of WO_X Films

XRD and XPS studies have been performed in this work, but many details of resistive switching in WO_X films remain to be resolved. Depth-profiling XPS is a common technique in understanding the stoichiometry and V_{OX} gradient in metal oxide films [2-3]. It has been reported that Ar milling process potentially reduces the WO_X film through ion bombardment and the dissimilar etch rates between W and O [4], so one needs to come up with an ingenious method to compensate for the contributions from these effects. Many other techniques can be potentially used in this kind of study, including atomic force microscopy (AFM) [5], transmission electron microscopy (TEM) [6], high-resolution transmission electron microscopy (HRTEM) [7], selected-area electron diffraction (SAED) [7], scanning tunneling microscopy (STM) [8], Rutherford backscattering spectrometry (RBS) [6], Raman spectroscopy [9], etc. These techniques will provide useful information on the structure, morphology, stoichiometry, and the conduction properties. Additionally, *in situ* electrical studies combined with material characterization techniques will be very useful in helping understand the switching dynamics [10].

8.2.2 Quantitative Study of Resistive Switching in WO_X memristors

In addition to material characterization, controlled electrical measurements on devices with designed parameters would also render much information on the resistive switching and transport mechanisms. For example, by manipulating the temperature and vacuum level during electrical testing, switching dynamics can be quantitatively analyzed;

and by changing physical dimensions of WO_X memristors, conduction mechanisms can be studied. The above testing results can be further used to verify and optimize the proposed phenomenological model, and may even provide sufficient evidence for an atomic-level memristor model.

8.2.3 Characterizations and Functionality of CMOS-Integrated Memristor Arrays

The CMOS-integrated memristor arrays were successfully fabricated, but characterizing and optimizing efforts are still required. As mentioned earlier, memristors are interesting for their ability to form large networks, so fabrication yield and switching uniformity are important factors that call for attentions. Once high-yield memristor networks have been fabricated, one needs to design appropriate CMOS circuits and algorithms to achieve functions. These networks could be used simply as analog memories, but could also be used as logic and signal processing units [1].

8.2.4 Relating Calcium Ion Dynamics with Oxygen Vacancy Dynamics

Extensive study in neuroscience has provided a unified picture on how calcium ions affect the synaptic plasticity and it is believed to initiate the release of neurotransmitters on the pre-synaptic end and the activation of receptors on the post-synaptic end [11-12]. Moreover, the transients of Ca^{2+} ion concentration act as an internal clock that regulates the different rate-dependent and spike-timing-dependent behaviors. Since calcium ions and oxygen vacancies are both charged particles that transport through a medium under the force of a potential difference, it is possible that the dynamics of both particles follow a similar transport behavior, and possibly at

comparable time scales. Therefore, if one can accurately map the dynamics of Ca^{2+} to those of V_{OX}, the synaptic effects observed in WO_X memristors can be unified using the same biological model so that the emulation of WO_X memristors as synapses become more realistic.

8.2.5 Interaction between Memristor Synapses and CMOS Neurons in a Network

Since synapses and neurons together make up neural networks, it is difficult to decouple the progression of each of them, especially in large, interconnected networks. It seems that now the community has started to gain sufficient knowledge about solid-state memristor synapses and CMOS neurons [13], and the hybrid memristor synapse/CMOS neuron system will provide a suitable platform for studying the interaction between synapses and neurons in a network, namely, how input patterns affect the network dynamics and how network dynamics affect the output. These findings will not only help us understand the operations of neural networks, either biological or memristor-based, but can also be applied to areas requiring "fuzzy logic" such as pattern recognition, financial forecast, medical diagnosis, data mining, and telecommunications [14].

References

[1] K. K. Likharev, "CrossNets: Neuromorphic Hybrid CMOS/Nanoelectronic Networks," *Sci. Adv. Mater.* **3**, 322, 2011.

[2] K. P. Biju, X. Liu, M. Siddik, S. Kim, J. Shin, I. Kim, A. Ignatiev, H. Huang, "Resistive Switching Characteristics and Mechanism of Thermally Grown WO_X Thin Films," *J. Appl. Phys.* **110**, 064505, 2011.

[3] O. Bohnke, G. Frand, M. Fromm, J. Weber, O. Greim, "Depth Profiling of W, O and H in Tungsten Trioxide Thin Films Using RBS and ERDA Techiniques," *Appl. Surf. Sci.* **93**, 45, 1996.

[4] T. J. Driscoll, L. D. McCormick, W. C. Lederer, “Altered Layer Formation and Sputtering Yields for 5 keV Ar^+ Bombardment of MoO_3 and WO_3,” *Surf. Sci.* **187**, 539, 1987.

[5] D. S. Shang, L. Shi, J. R. Sun, B. G. Shen, F. Zhuge, R. W. Li, Y. G. Zhao, “Improvement of Reproducible Resistance Switching in Polycrystalline Tungsten Oxide Films by *in situ* Oxygen Annealing,” *Appl. Phys. Lett.* **96**, 072103, 2010.

[6] J. P. Strachan, M. D. Pickett, J. J. Yang, S. Aloni, A. L. D. Kilcoyne, G. Medeiros-Ribeiro, R. S. Williams, “Direct Identification of the Conducting Channels in a Functioning Memristive Device,” *Adv. Mater.* **22**, 3573, 2010.

[7] D. H. Kwon, K. M. Kim, J. H. Jang, J. M. Jeon, M. H. Lee, G. H. Kim, X. S. Li, G. S. Park, B. Lee, S. Han, M. Kim, C. S Hwang, “Atomic Structure of Conducting Nanofilaments in TiO_2 Resistive Switching Memory, “ *Nat. Nanotechnol.* **5**, 148, 2010.

[8] S. Santucci, C. Cantalini, M. Crivellari, L. Lozzi, L. Ottaviano, M. Passacantando, “X-Ray Photoemission Spectroscopy and Scanning Tunneling Spectroscopy Study on the Thermal Stability of WO_3 Thin Films,” *J. Vac. Sci. Technol. A* **18**, 1077, 2000.

[9] C. Bittencourt, R. Landers, E. Llobet, X. Correig, J. Calderer, “The Role of Oxygen Partial Pressure and Annealing Temperature on the Formation of W=O Bonds in Thin WO_3 Films,” *Semicond. Sci. Technol.* **17**, 522 (2002).

[10] Y. Yang, P. Gao, S. Gaba, T. Chang, X. Pan, W. Lu, “Observation of Conducting Filament Growth in Nanoscale Resistive Memories,” *Nat. Commun.* **3**, 732, 2012.

[11] H. Z. Shouval, M. F. Bear, L. N. Cooper, “A Unified Model of NMDA Receptor-Dependent Bidirectional Synaptic Plasticity,” *Proc. Natl. Acad. Sci. USA* **99**, 10831, 2002.

[12] H. Z. Shouval, G. C. Castellani, B. S. Blais, L. C. Yeung, L. N. Cooper, “Converging Evidence for a Simplified Biophysical Model of Synaptic Plasticity,” *Biol. Cybern.* **87**, 383, 2002.

[13] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S. C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, K. Boahen, “Neuromorphic Silicon Neuron Circuits,” *Front. Neurosci.* **5**, 73, 2011.

[14] B. Widrow, D. E. Rumelhart, M. A. Lehr, “Neural Networks: Applications in Industry, Business and Science,” *Commun. ACM* **37**, 93, 1994.